

This Page Is Inserted by IFW Operations
and is not a part of the Official Record

BEST AVAILABLE IMAGES

Defective images within this document are accurate representations of the original documents submitted by the applicant.

Defects in the images may include (but are not limited to):

- BLACK BORDERS
 - TEXT CUT OFF AT TOP, BOTTOM OR SIDES
 - FADED TEXT
 - ILLEGIBLE TEXT
 - SKEWED/SLANTED IMAGES
 - COLORED PHOTOS
 - BLACK OR VERY BLACK AND WHITE DARK PHOTOS
-
- GRAY SCALE DOCUMENTS

IMAGES ARE BEST AVAILABLE COPY.

**As rescanning documents *will not* correct images,
please do not report the images to the
Image Problem Mailbox.**

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 13:48:40 ; Search time 1862.98 seconds
(without alignments)
1495.893 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
1 atgaaataatgacaaatcg.....tagctacagaatcaataa 621

Scoring table: IDENTITY_NUC
Gapox 10.0 , Gapext 1.0

Searched: 12496460 seqs, 2243815623 residues

Total number of hits satisfying chosen parameters: 24992920

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 10%
Listing first 45 summaries

Database :

Pending_Patents_NA:*
1: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
2: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
3: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
4: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
5: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
6: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
7: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
8: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
9: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
10: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
11: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
12: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
13: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
14: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
15: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
16: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
17: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
18: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
19: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
20: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
21: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
22: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
23: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
24: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
25: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
26: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
27: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
28: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
29: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
30: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
31: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
32: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
33: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
34: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
35: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
36: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
37: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
38: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
39: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
40: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
41: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
42: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*
43: /cgn2_6/ptodata/1/pna/US0877.COMB.seq:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	621	100.0	621	US-09-292-437-2	Sequence 2, Appli
2	621	100.0	3733	US-08-781-986A-538	Sequence 538, App
3	621	100.0	3733	US-08-956-171-538	Sequence 538, App
4	621	100.0	3733	US-08-956-171B-538	Sequence 538, App
5	611.4	98.5	624	US-08-827-356-2678	Sequence 2678, Ap
6	611.4	98.5	624	US-09-611-529-2938	Sequence 2938, Ap
7	611.4	98.5	91077	US-08-831-156A-108	Sequence 108, App
8	611.4	98.5	812842	US-09-611-529-7449	Sequence 7449, App
9	572.2	92.1	45487	US-06-038-697-735	Sequence 735, App
10	572.2	92.1	45488	US-06-046-714-834	Sequence 834, App
11	534.2	86.0	551	US-09-620-608-849	Sequence 849, App
12	534.2	86.0	551	US-06-144-883-849	Sequence 849, App
13	340.2	54.8	612	US-09-450-969-1132	Sequence 1132, Ap
14	331.4	53.4	11428	US-06-068-628-360	Sequence 360, App
15	331.4	53.4	11434	US-06-038-081-348	Sequence 348, App
16	331.4	53.4	11434	US-06-046-933-355	Sequence 355, App
17	194	31.2	310	US-06-206-848-39	Sequence 39, Appl
18	194	31.2	310	US-06-242-578-165	Sequence 165, App
19	194	31.2	310	US-06-257-931-163	Sequence 163, App
20	194	31.2	310	US-06-253-625-165	Sequence 165, App
21	194	31.2	310	US-06-269-308-165	Sequence 165, App
22	75.2	12.1	1517	US-09-404-520-13349	Sequence 13349, A
23	69.2	11.1	641	US-06-211-750-24644	Sequence 24644, A
24	68.8	11.1	786	US-09-397-761A-2720	Sequence 2720, Ap
25	68.8	11.1	786	US-09-641-529-376	Sequence 376, App
26	68.4	11.0	599	US-06-207-458-143594	Sequence 143594, A
27	68.2	11.0	889	US-09-397-761A-2719	Sequence 2719, Ap
28	68.2	11.0	889	US-09-641-529-411	Sequence 411, App
29	66.6	10.7	666	US-09-397-761A-2717	Sequence 2717, Ap
30	66.6	10.7	666	US-09-641-529-380	Sequence 380, App
31	66.2	10.7	585	US-06-211-750-11125	Sequence 11125, A
32	66	10.6	466	US-09-565-306-29077	Sequence 29077, A
33	66	10.6	466	US-06-207-458-58087	Sequence 58087, A
34	66	10.6	867	US-09-397-761A-2710	Sequence 2710, Ap
35	66	10.6	867	US-09-641-529-448	Sequence 448, App
36	65.6	10.6	666	US-06-207-458-88810	Sequence 88810, A
37	65.6	10.6	799	US-09-404-520-13067	Sequence 13067, A
38	65.4	10.5	7218	US-08-466-194-14	Sequence 14, Appl
39	65	10.5	972	US-09-397-761A-2715	Sequence 2715, Ap
40	65	10.5	972	US-09-641-529-680	Sequence 680, App
41	65	10.5	4587	US-06-212-659-6	Sequence 6, Appli
42	65	10.5	32768	US-06-208-020-72	Sequence 72, Appl
43	65	10.5	32768	US-06-209-043-95	Sequence 95, Appl
44	65	10.5	32768	US-06-213-172-35	Sequence 35, Appl
45	65	10.5	32768	US-06-213-173-35	Sequence 35, Appl

ALIGNMENTS

```
RESULT 1
US-09-292-437-2
; Sequence 2, Application US/09292437
; GENERAL INFORMATION:
; APPLICANT: Olaf Schneewind
; APPLICANT: Sarkis Mazmanian
; APPLICANT: Gwen Liu
; APPLICANT: Hung Ton-That
; FILE OF INVENTION: IDENTIFICATION OF SORTASE GENE
; FILE REFERENCE: 510015.213
; CURRENT APPLICATION NUMBER: US/09/292.437
; CURRENT FILING DATE: 1999-04-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 2
; LENGTH: 621
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
US-09-292-437-2
```

```
Query Match 100.0%; Score 621; DB 55; Length 621;
Best Local Similarity 100.0%; Pred. No. 9.8e-107;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaaaaatgacaaatcgattatgacaaatcgctggtggtgactatctagtgca 60
Db 1 atgaaaaatgacaaatcgattatgacaaatcgctggtggtgactatctagtgca 60
QY 61 gcatattgttctgtaaacacacatcgataatcttcaacgtaagaataagatgaa 120
Db 61 gcatattgttctgtaaacacacatcgataatcttcaacgtaagaataagatgaa 120
QY 121 aagattgaacacatgataaaatgttaaaagacagcgagtaagaataaaacagcaa 180
Db 121 aagattgaacacatgataaaatgttaaaagacagcgagtaagaataaaacagcaa 180
QY 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
Db 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
QY 241 gctgatatataagacagcatatccagacacacacacacacacacacacacacac 300
Db 241 gctgatatataagacagcatatccagacacacacacacacacacacacacacac 300
QY 301 gtaagcttcgaaagaataatgatactatgataatgataatgataatgataatgata 360
Db 301 gtaagcttcgaaagaataatgatactatgataatgataatgataatgataatgata 360
QY 361 acttcattgacgctccgaactatcaatttacaatcttaagaacagcaaaaagatgagt 420
Db 361 acttcattgacgctccgaactatcaatttacaatcttaagaacagcaaaaagatgagt 420
QY 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgataagatgataagat 480
Db 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgataagatgataagat 480
QY 481 gtttaagctcagatgtaggagttctagatgaaacaaaagtaagaataaactaacaa 540
Db 481 gtttaagctcagatgtaggagttctagatgaaacaaaagtaagaataaactaacaa 540
QY 541 ttaattactgtatgataatcaatgaaagacagcgcttgggaaaaaacgtaaatcttt 600
Db 541 ttaattactgtatgataatcaatgaaagacagcgcttgggaaaaaacgtaaatcttt 600
QY 601 gtagctacagaagtcacaataa 621
Db 601 gtagctacagaagtcacaataa 621
```

```
RESULT 2
US-08-781-986A-538/c
; Sequence 538, Application US/08781986A
```

```
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/781, 986A
; FILING DATE:
; CLASSIFICATION: 435
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248PP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 538:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-781-986A-538
```

```
Query Match 100.0%; Score 621; DB 11; Length 3733;
Best Local Similarity 100.0%; Pred. No. 1.2e-106;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

QY 1 atgaaaaatgacaaatcgattatgacaaatcgctggtggtgactatctagtgca 60
Db 730 atgaaaaatgacaaatcgattatgacaaatcgctggtggtgactatctagtgca 671
QY 61 gcatattgttctgtaaacacacatcgataatcttcaacgtaagaataagatgaa 120
Db 670 gcatattgttctgtaaacacacatcgataatcttcaacgtaagaataagatgaa 611
QY 121 aagattgaacacatgataaaatgttaaaagacagcgagtaagaataaaacagcaa 180
Db 610 aagattgaacacatgataaaatgttaaaagacagcgagtaagaataaaacagcaa 551
QY 181 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 240
Db 550 gctaaacctcaattccgaaagataatcgaaagtcgacgctatctgtaaatccagat 491
QY 241 gctgatatataagacagcatatccagacacacacacacacacacacacacacac 300
Db 490 gctgatatataagacagcatatccagacacacacacacacacacacacacacac 431
QY 301 gtaagcttcgaaagaataatgatactatgataatgataatgataatgataatgata 360
Db 430 gtaagcttcgaaagaataatgatactatgataatgataatgataatgataatgata 371
QY 361 acttcattgacgctccgaactatcaatttacaatcttaagaacagcaaaaagatgagt 420
Db 370 acttcattgacgctccgaactatcaatttacaatcttaagaacagcaaaaagatgagt 311
QY 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgataagatgataagat 480
Db 421 atgggtgactttaaagtctgtaagaacacgtaagataataatgataagatgataagat 480
```


Db 310 ATGCTGCTCTTAAGTGTGTAATGAAACACGTAAGTAAATGACAAGTAAAGAGAT 251
Qy 481 gtaagcctacagatgtagaggttcagatgacacaaagtaaaagataaacaatlaaca 540
Db 250 GTTAAGCCTACAGATGTAGAGTCTTAGATGACAAAAGTAAGATTAACATTAACA 191
Qy 541 ttaattactgtgtagatgatacaatgaaagacagcgcttgggaaacgttaaatcttt 600
Db 190 TTAATTACTGTGATGATTAACATGAAAAGACAGCGCTTGGGAAAACGTAAATCTTT 131
Qy 601 gtaagctacagaaatgcaataa 621
Db 130 GTAGCTACAGAAGTCAATAA 110

RESULT 3

US-08-956-171-538/c
; Sequence 538, Application US/08956171
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.171
; FILING DATE:
; CLASSIFICATION: 424
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER:
; FILING DATE:
; ATTORNEY/AGENT INFORMATION:
; NAME: Benson, Bob
; REGISTRATION NUMBER: 30,446
; REFERENCE/DOCKET NUMBER: PB248BP
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 309-8504
; TELEFAX: (301) 309-8512
; INFORMATION FOR SEQ ID NO: 538:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 3733 base pairs
; TYPE: nucleic acid
; STRANDEDNESS: double
; TOPOLOGY: linear
US-08-956-171-538

Query Match 100.0%; Score 621; DB 13; Length 3733;

Best Local Similarity 100.0%; Pred. No. 1.2e-106; Mismatches 0; Indels 0; Gaps 0;

Qy 1 atgaaaaatgacaaatcgatgaaatgacaaatcgctgtgtgtgtactatcctagtgaca 60
Db 730 ATGAAAAATGACAAATCGATTAATGCAATCGCTGTGTGTGTACTTCTTCTAGTGCA 671
Qy 61 gcatattgttctgtaaacacataatcgaataatcttcacgataaagataaagatgaa 120
Db 670 GCATATTGTGTTGCTAAACACATATCATATATATCTTACACGATAAAGATTAAGATGAA 611
Qy 121 aagatgacaaatgataaataatgtaaaagacagcgagatgaaagataaagacagcaa 180
Db 610 AAGATTGAACATATGATTAATAATGTAAAGACAGCGCAGTAAGATTAATAAGACAGCAA 551

Qy 181 gctaaacctcaaatccgaaagataaalcgaaagtggcaggtatattgaaatccagat 240
Db 550 GCTAAACCTCAAAATCCGAAGATTAATCGAAAAGTGCGAGGCTATATTGAATTCACAT 491
Qy 241 gctgatatgaaagaccagatatatccagagcagcaaccctggaacaatlaaatgaggt 300
Db 490 GCTGATATTAAAGAACCACTATATTCAGAGCAGCAACACCTGAACAAATTAATAGAGGT 431
Qy 301 gtaagcttgcagaagaaatgaatcactagatgatacaaatatltcaatltgcaagac 360
Db 430 GTAGCTTGCGAGAGAAAATGATCATGATGATCAAAATATTTCATTTCAGAGACAC 371
Qy 361 acttcaatgacgcgtccgaactatcaattacaactttaaagcagccaaaaaggtagt 420
Db 370 ACTTCATGTGACGCCGCCAATCATATTTACAAATCTTAAGACACCAAAAAGTAGT 311
Qy 421 atggtgtaactttaaagttggtatgaaacagtaagataaataatgcaagataaagat 480
Db 310 ATGCTGCTCTTAAGTGTGTAATGAAACACGTAAGTAAATGACAAGTAAAGAGAT 251
Qy 481 gtaagcctacagatgtagaggttcagatgacacaaagtaaaagataaacaatlaaca 540
Db 250 GTTAAGCCTACAGATGTAGAGTCTTAGATGACAAAAGTAAGATTAACATTAACA 191
Qy 541 ttaattactgtgtagatgatacaatgaaagacagcgcttgggaaacgtlaaatcttt 600
Db 190 TTAATTACTGTGATGATTAACATGAAAAGACAGCGCTTGGGAAAACGTAAATCTTT 131
Qy 601 gtaagctacagaaatgcaataa 621
Db 130 GTAGCTACAGAAGTCAATAA 110

RESULT 4

US-08-956-171B-538/c
; Sequence 538, Application US/08956171B
; GENERAL INFORMATION:
; APPLICANT: Charles Kunsch
; GIL H. Choi
; Patrick S. Dillon
; Craig A. Rosen
; Steven C. Barash
; Michael R. Fannon
; TITLE OF INVENTION: Staphylococcus aureus Polynucleotides and Sequences
; NUMBER OF SEQUENCES: 5255
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: Human Genome Sciences, Inc.
; STREET: 9410 Key West Avenue
; CITY: Rockville
; STATE: Maryland
; COUNTRY: USA
; ZIP: 20850
; COMPUTER READABLE FORM:
; MEDIUM TYPE: Diskette, 3.50 inch, 1.4MB storage
; COMPUTER: HP Vectra 486/33
; OPERATING SYSTEM: MSDOS version 6.2
; SOFTWARE: ASCII Text
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/08/956.171B
; FILING DATE: 20-Oct-1997
; CLASSIFICATION: <Unknown>
; PRIOR APPLICATION NUMBER:
; APPLICATION NUMBER: 60/009 861
; FILING DATE: January 5, 1996
; APPLICATION NUMBER: 08/781,986
; FILING DATE: January 3, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Hoover, Kenley K.
; REGISTRATION NUMBER: 40,302
; REFERENCE/DOCKET NUMBER: PB248P1
; TELECOMMUNICATION INFORMATION:
; TELEPHONE: (301) 610-5790
; TELEFAX: (301) 309-8439

```

: INFORMATION FOR SEQ ID NO: 538:
:   SEQUENCE CHARACTERISTICS:
:     LENGTH: 3733 base pairs
:     TYPE: nucleic acid
:     STRANDEDNESS: double
:     TOPOLOGY: linear
:   SEQUENCE DESCRIPTION: SEQ ID NO: 538:
US-08-956-1718-538

```

```

Query Match      100.0%: Score 621: DB 13: Length 3733:
Best Local Similarity 100.0%: Pred. No. 1.2e-106:
Matches 621: Conservative 0: Mismatches 0: Indels 0: Gaps 0:

```

```

QY 1 atgaaaaaagagacaaatcgatlaaagcaatcgctggtggtgactatccctagtgca 60
   |||||||
DB 730 ATGAAAAAATGGACAAATCGATTATATGCAATCGCTGCTGGTACTTATCTAGTGCA 671
   |||||||
QY 61 gcatattgttgcctaaaccacatatacttaattccttcacgaataaagaatgaa 120
   |||||||
DB 670 GCATATTGTTGCTTAACCCATATCGATATATCTTCCGATTAAGATAAGATGAA 611
   |||||||
QY 121 aaagttgaacaatatgataaaatgtaaaagaaagcagcgtaaaagtaaaagcagaa 180
   |||||||
DB 610 AAGATTGAACAATATGATAAAATGTAAGAACAGCGCATAGATAAAGACACGCAA 551
   |||||||
QY 181 gctaaacctcaaatccgaaagataatcgaaagtgacagctatattgaaatccagat 240
   |||||||
DB 550 GCTTAACCTCAATTCGAAAGATTAATCGAAGTGGCAGGCTATTTGAATTCAGAT 491
   |||||||
QY 241 gctgataataaagacagatataccagagcagcaacacctgaacaattaaatagagt 300
   |||||||
DB 490 GCTGATATTAAAGACCGATATATCCAGACACACCTGAACATTAATATGAGGT 431
   |||||||
QY 301 gtaagcttgcagaagaagaatgtaactgatatgataaaatatttcaattcgagagac 360
   |||||||
DB 430 GTAAGCTTTCCAGAGAAATGAAATCACTAGATGATCAAAATATTTCATATTCAGAGAC 371
   |||||||
QY 361 acttcaatgacgcctcgaactatcaatcttaaatcgaagcagcaaaaagtgagt 420
   |||||||
DB 370 ACTTCAATGACGCTCGAAGCTATTCATTTAACAATCTTAAGACGCAAAAAGGATGT 311
   |||||||
QY 421 atgggtactttaaaagttgtaataagaaacgtaagtataaaatgcaagataagagt 480
   |||||||
DB 310 ATGGGTACTTTAAAGTTGGTAATGAACACGTAAAGTATAAATGACAAGATTAAGAGAT 251
   |||||||
QY 481 gtaagacctacagatgtaggaagttctgatatgaacaaaaggtaaaagtaaaacaa 540
   |||||||
DB 250 GTTAAGCCTACAGATGTAGAGGTTCTAGATGAAACAAAAGGTAAGATTAACA 191
   |||||||
QY 541 ttaactactgtgatagttaacaatgaagaacagcgcttgagaaaaacgttaaatctt 600
   |||||||
DB 190 TTATATTCTTGATGTATTCATATGAAGAACAGACGCGCTTGGGAAAAACGTAATCTTT 131
   |||||||
QY 601 gtagctacagaagtcataaa 621
   |||||||
DB 130 GTAGCTACAGAACTCAATAA 110

```

```

RESULT 5
US-08-827-356-2678
: Sequence 2678, Application US/08827356
: GENERAL INFORMATION:
:   APPLICANT: George H. Shimer, Jr.
:   APPLICANT: Robert S. Hare
:   APPLICANT: Karen J. Shaw
:   TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
:   NUMBER OF SEQUENCES: 5574
:   CORRESPONDENCE ADDRESS:
:   ADDRESSEE: Schering-Plough Corporation
:   STREET: 2000 Galloping Hill Road

```

```

: CITY: Kenilworth
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07033-0530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: IBM, PC compatible
: SOFTWARE: Patent
: CURRENT APPLICATION DATA:
:   APPLICATION NUMBER: US/08/827,356
:   FILING DATE: 01-APR-1997
:   PRIOR APPLICATION DATA:
:     APPLICATION NUMBER: 60/014,477
:     FILING DATE: 01-APR-1996
:     APPLICATION NUMBER: 60/016,743
:     FILING DATE: 02-MAY-1996
:     APPLICATION NUMBER: 60/020,016
:     FILING DATE: 14-JUN-1996
:   INFORMATION FOR SEQ ID NO: 2678:
:     SEQUENCE CHARACTERISTICS:
:       LENGTH: 624 base pairs
:       TYPE: nucleic acid
:       STRANDEDNESS: double
:       TOPOLOGY: circular
:     MOLECULE TYPE: DNA (genomic)
:     HYPOTHETICAL: NO
:     ANTI-SENSE: NO
:     ORIGINAL SOURCE:
:       ORGANISM: Staphylococcus aureus
:     FEATURE:
:       NAME/KEY: misc.feature
:       LOCATION: 1..624
US-08-827-356-2678

```

```

Query Match      98.5%: Score 611.4: DB 12: Length 624:
Best Local Similarity 99.0%: Pred. No. 6.1e-105:
Matches 615: Conservative 0: Mismatches 6: Indels 0: Gaps 0:

```

```

QY 1 atgaaaaaagagacaaatcgatlaaagcaatcgctggtggtgactatccctagtgca 60
   |||||||
DB 4 ATGAAAAAATGGACAAATCGATTATATGCAATCGCTGCTGGTACTTATCTAGTGCA 63
   |||||||
QY 61 gcatattgttgcctaaaccacatatacttaattccttcacgaataaagaatgaa 120
   |||||||
DB 64 GCATATTGTTGCTTAACCCATATCGATATATCTTCCGATTAAGATTAAGATGAA 123
   |||||||
QY 121 aaagttgaacaatatgataaaatgtaaaagaaagcagcgtaaaagataaaagcagaa 180
   |||||||
DB 124 AAGATTGAACAATATGATAAAATGTAAGAACAGCGGAGTAAGAACATTAAGACGCAA 183
   |||||||
QY 181 gctaaacctcaaatccgaaagataaatcgaaagtgacaggtatataatgaaatccagat 240
   |||||||
DB 184 GCTAAACCTCAAAATTCGAAAGATTAATCAAAAGGCGAGGCTATTTGAATTCAGAT 243
   |||||||
QY 241 gctatatttaagaacacagatataccagagcagcaaaccttgaaacaattaaatagagt 300
   |||||||
DB 244 GCTGATATTAAAGAACCGATATATCCAGACGACGAAACCTGAAACATTTAATAGAGT 303
   |||||||
QY 301 gtaagcttgcagaagaagaatgtaactacatagatgatacaaaatattcaattgcagacac 360
   |||||||
DB 304 GTAAGCTTTGCAGAAAGAAATGAAATCACTAGATGATCAAAATATTTCATATTCAGAGAC 363
   |||||||
QY 361 acttcaatgacgcctcgaactatcaatctcaaatcctaagcagccaaaaagtgagt 420
   |||||||
DB 364 ACTTCAATGACCGCTCGAATCAATCAATTTAACAATCTTAAGACAGCCAAAAGGTAGT 423
   |||||||
QY 421 atgggtactttaaaagttgtaataagaaacgctgaagtataaaatgcaagataagagt 480
   |||||||
DB 424 ATGGGTACTTTAAAGTTGGTAATGAACACGTAAAGTATAAATGACAAGATTAAGAGAT 483
   |||||||
QY 481 gtaagacctacagatgtaggaagttctgatatgaacaaaaggtaaaagtaaaacaa 540

```

```

Db      484  ttaagccacagatgtatagacgtttatgatgaccaaaagstaaagattaaacatttaca 543
Oy      541  ttaattactgtgatgatattacaatgtaaaaagacagcgcttltgggaaaaacgtlaaaatttc 600
Db      544  ttaattactgtgatgatattacaatgtaaaaagacagcgcttltgggaaaaacgtlaaaatttc 603
Oy      601  gtacgtacagaagtcacaaata 621
Db      604  gtacgtacagaagtcacaaata 624

RESULT 6
US-09-611-529-2938
Sequence 2938, Application US/09611529
GENERAL INFORMATION:
APPLICANT: George H. Shimer, Jr.
APPLICANT: George H. Miller
APPLICANT: Roberta S. Hare
APPLICANT: Karen J. Shaw
TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
FILE REFERENCE: 1034/1C963J051
CURRENT APPLICATION NUMBER: US/09/611,529
CURRENT FILING DATE: 2000-06-30
PRIORITY APPLICATION NUMBER: US 09/417,811
PRIORITY FILING DATE: 1999-10-14
PRIORITY APPLICATION NUMBER: US 09/353,718
PRIORITY FILING DATE: 1999-07-14
PRIORITY APPLICATION NUMBER: US 09/266,557
PRIORITY FILING DATE: 1999-03-11
PRIORITY APPLICATION NUMBER: US 09/266,556
PRIORITY FILING DATE: 1999-03-11
PRIORITY APPLICATION NUMBER: US 09/266,555
PRIORITY FILING DATE: 1999-03-11
PRIORITY APPLICATION NUMBER: US 09/266,542
PRIORITY FILING DATE: 1999-03-11
PRIORITY APPLICATION NUMBER: US 09/266,541
PRIORITY FILING DATE: 1999-03-11
PRIORITY APPLICATION NUMBER: US 09/037,934
PRIORITY FILING DATE: 1998-03-10
PRIORITY APPLICATION NUMBER: US 09/036,720
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,338
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,334
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,221
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,137
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,082
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,081
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/036,079
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/035,913
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 09/035,744
PRIORITY FILING DATE: 1998-03-06
PRIORITY APPLICATION NUMBER: US 08/827,356
PRIORITY FILING DATE: 1997-04-01
PRIORITY APPLICATION NUMBER: US 08/831,156
PRIORITY FILING DATE: 1997-04-01
PRIORITY APPLICATION NUMBER: US 60/014,477
PRIORITY FILING DATE: 1996-04-01
PRIORITY APPLICATION NUMBER: US 60/016,743
PRIORITY FILING DATE: 1996-05-02
PRIORITY APPLICATION NUMBER: US 60/020,016
PRIORITY FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 7451
SEQ ID NO 2938
LENGTH: 624

```

```

: TYPE: DNA
: ORGANISM: Staphylococcus aureus
US-09-611-529-2938

Query Match          98.5%: Score 611.4: DB 22: Length 624;
Best Local Similarity 99.0%: Pred No. 6.1e-105;
Matches 615: Conservative 0: Mismatches 6: Indels 0: Gaps 0:

OY 1 atgaaaaaatggacaatcgcattaaatgacaatcgctgtgtgtgtacttaacctagtggca 60
Db 4 atgaaaaaatggacaatcgcattaaatgacaatcgctgtgtgtgtacttaacctagtggca 63
OY 61 gcatatttgtttgtgtaaacacatctgataattcttcacgataaagaataaagtga 120
Db 64 gcatatttgtttgtgtaaacacacatctgataattcttcacgataaagaataaagtga 123
OY 121 aagattgacacatatgataaanaatgtaaaagacagcgagtaagaagataaaagcagca 180
Db 124 aagattgacacatatgataaanaatgtaaaagacagcgagtaagaagataaaagcagca 183
OY 181 gctaaacctcaaatccgaaagaataatcogaagtgycgcggtctatatctgaattccagat 240
Db 184 gctaaacctcaaatccgaaagaataatcogaagtgycgcggtctatatctgaattccagat 243
OY 241 gctgatatlaaagaaccagtatatccagagaccagcaacacactgaacaattaaatagagat 300
Db 244 gctgatatlaaagaaccagtatatccagagaccagcaacacactgaacaattaaatagagat 303
OY 301 gtaagctttgcagaagaanaatgtaatcactagatgataatattccaatttcacgagcac 360
Db 304 gtaagctttgcagaagaanaatgtaatcactagatgataatattccaatttcacgagcac 363
OY 361 aacttcattgacgcgcggaactcaatcttaacaaacttaagcgccaaaaagtgagt 420
Db 364 aacttcattgacgcgcggaactcaatcttaacaaacttaagcgccaaaaagtgagt 423
OY 421 atggtgtactttaaagtttgtgtaatgaaacacgtaagataaaatgacaagtataagagat 480
Db 424 atggtgtactttaaagtttgtgtaatgaaacacgtaagataaaatgacaagtataagagat 483
OY 481 gttaaagctacaagatgtagggagttcttagatgacaacaaaaaggtaaagatataacaat 540
Db 484 gttaaagccacaagatgtaggaagttcttagatgacaacaaaaaggtaaagatataacaat 543
OY 541 ttaattactttgattgattacaatgaaagacagcgctttgggaaaaacgtataaattctt 600
Db 544 ttaattactttgattgattacaatgaaagacagcgctttgggaaaaacgtataaattctt 603
OY 601 gttagctacagaagtcaataaa 621
Db 604 gttagctacagaagtcaataaa 624

RESULT 7
US-08-831-156A-108
: Sequence 108, Application US/08831156A
: GENERAL INFORMATION:
: APPLICANT: George H. Shimer, Jr.
: APPLICANT: George H. Miller
: APPLICANT: Robert S. Hare
: APPLICANT: Karen J. Shaw
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCE OF A
: TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS
: TITLE OF INVENTION: GENOME AND RELATED
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: NUMBER OF SEQUENCES: 112
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Schering-Plough Corporation
: STREET: 2000 Galloping Hill Road
: CITY: Kenilworth
: STATE: New Jersey
: COUNTRY: USA

```

```
ZIP: 07033-0530
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/831,156A
FILING DATE: 01-APR-1997
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 60/014,477
FILING DATE: 01-APR-1996
APPLICATION NUMBER: 60/016,743
FILING DATE: 02-MAY-1996
APPLICATION NUMBER: 60/020,016
FILING DATE: 14-JUN-1996
INFORMATION FOR SEQ ID NO: 108:
SEQUENCE CHARACTERISTICS:
LENGTH: 91077 base pairs
TYPE: nucleic acid
STRANDEDNESS: double
TOPOLOGY: circular
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY: misc_feature
LOCATION: 1...91077
US-08-831-156A-108
```

```
Query Match          98.5%  Score 611.4:  DB 12:  Length 91077:
Best Local Similarity 99.0%:  Pred. No 1.2e-104:
Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0;

QY 1  atgaaaaatgagcaaatcgataatgacaatcgctgtgtgtgactatcctagtggca 60
Db 15161 ATGAAAAATGAGCAAAATCGATTATGATGACATCGCTGTGTGACTTATCTGAGTGGCA 15220

QY 61 gcatattgttgcataaccacacataatccttcacagataagaataagaatgaa 120
Db 15221 GCATATTGTTGCTTAACCAACATATGATTAATTTCTTCACGATTAAGATAAGATGAA 15280

QY 121 aaatgacaacataatgataaataatgataaagaacagcgagtaagaataagaacagcaa 180
Db 15281 AAGATTGAACAATRTGTATTAATAATGTTAAAGACAGCGAGTAAAGACAATTAACAGCAA 15340

QY 181 gctaaacctcaattccgaaagataatcgaaagtgacaggtatattgaaattccagat 240
Db 15341 GCTAAACCTCAAAATTCGAAAGATAATCAAAAGTGCAGCGCTATTGTAATTCAGAT 15400

QY 241 gctgatatataagaacagcatatccagagccagcaacactgtgaacataataagagat 300
Db 15401 GCTGATATTAAAGACCGATATATCCAGACCAACCTGGAACATTAATAAGAGAT 15460

QY 301 gtaagcttgcagaagaataatgaatcactagatgatacaaaatatttcattgcagagac 360
Db 15461 GTAAGCTTTCAGAGAAAGATGATCACTAGATGATCAAAAATATTTCATTTGCGAGCAC 15520

QY 361 acctcatgacgctccgaacatcaattacaattttaaagcagcgaacaaagagtagt 420
Db 15521 ACTTCAATGACCGTCGAGCATATCAATTTAACAATCTTAAAGCAGCAAAAAGAGTAGT 15580

QY 421 atgggtgactttaaagtgtgataatgaacacgtaagataataaatgacagataagagat 480
Db 15581 ATGGTGTACTTTAAAGTTGCTAATGAACACGTAAGTATTAATAAGACAGATATTAAGAGAT 15640

QY 481 gtttaagcctcaagatgtagagatctagatgaaacaaaagtaagaataagaataaacttaca 540
Db 15641 GTTAAGCCAACAGTGTAGAGATTCTAGATGAACAAAAGGTAAGATTAACAATTAACA 15700
```

```
QY 541 ttaattactgtatgattacaatgaaagacagcgcttgggaaaaacglaaaactctt 600
Db 15701 TTAATTACTGTGATGATTCAATGAAGACAGCGCTTTGGAAAAACGTAATAATCTTT 15760

QY 601 gtaactacagaagtcacaataa 621
Db 15761 GTAGCTACAGAAGTCACAATAA 15781

RESULT 8
US-09-611-529-7449/c
GENERAL INFORMATION:
APPLICANT: George H. Shimer, Jr.
APPLICANT: George H. Miller
APPLICANT: Roberta S. Hare
APPLICANT: Karen J. Shaw
TITLE OR INVENTION: Staphylococcus aureus Related Compositions and Methods
FILE REFERENCE: 1034/1C9630S1
CURRENT APPLICATION NUMBER: US/09/611,529
PRIOR FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/417,811
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/353,718
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 09/266,557
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,556
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,555
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,542
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,541
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/037,934
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: US 09/036,720
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,338
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,334
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,221
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,137
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,082
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,081
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,079
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/035,913
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/035,744
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 08/827,356
PRIOR FILING DATE: 1997-04-01
PRIOR APPLICATION NUMBER: US 08/831,156
PRIOR FILING DATE: 1997-04-01
PRIOR APPLICATION NUMBER: US 60/014,477
PRIOR FILING DATE: 1996-04-01
PRIOR APPLICATION NUMBER: US 60/016,743
PRIOR FILING DATE: 1996-05-02
PRIOR APPLICATION NUMBER: US 60/020,016
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 7451
SEQ ID NO 7449
LENGTH: 812842
TYPE: DNA
ORGANISM: Staphylococcus aureus
US-09-611-529-7449
```

```

Query Match: 98.5%; Score 611.4; DB 22; Length 812842;
Best Local Similarity 99.0%; Pred. No. 1.5e-104;
Matches 615; Conservative 0; Mismatches 6; Indels 0; Gaps 0

QY 1 atgaaataatggacaatcgatataatgacaatcgcgtgtgtgtaacttccagtgtgca 60
Db 362980 ATGAAATAATGACAATTCGATTTAATGACAATCGCGTGTGTAGTACTTATCCTAGTGCA 362921

QY 61 gcatattgtttgtccaaaccacatatcgataattatcttccagataaagaatgtga 120
Db 362920 GCATTTTGTGTTCGTAACACCATATTCGATAATTATCTTCACGATTAACATAAAGATGAA 362861

QY 121 aagatgtgacaaatcgtataaataatgaaagacagcgcgatgaagataaaagcagcaa 180
Db 362860 AAGATTTGAACAATATGATATTAATAATGTTAAAGAACACGCGCATTAACACATTAAGCAGCAA 362801

QY 181 gctaaacctcaaatctcgaagaataaattcgaagttgacgcbatatatgaaattccagat 240
Db 362800 GCTAAACCTCAAAATTCGGAAGATATTAATCAAAAGTGCGCAGGCTATATTCAAATTCAGAT 362741

QY 241 gctgtattaaagaacccagatatatccagacagcaacccctgaacatataaagaagt 300
Db 362740 GCTGTATTTAAAGAACCGTATATTCACGAGACCAACGACCTGAAACAATTTAATAGAGGT 362681

QY 301 gtaagctcttcgagaagaataatgatacctcaga tgaatcaaaatattcgaattgcagacac 360
Db 362680 GTAACCTTTGCAGAGAAGAAATAATACATCAGTAGATCAAAATATTTCAATTTGCAGACAC 362621

QY 361 acttcattgacgcgtccgacatactcaatttcaaaattcctaagcagcaaaaaagtgagt 420
Db 362620 ACTTTCATTGTGACGCGTCGACCATATCAATTTACAAATCTTTAAAGCACCAAAAAAGGTAGT 362561

QY 421 atggtgtaacctaaagttcgtgtaatgaacaacgtaagtatataaaatgacaagtataagagt 480
Db 362560 ATGCGTACTTAAAGTTGATATGATAACACGTAAGTATATAAATGCAAGTATTAAGAGAT 362501

QY 481 gttaaagctacagatgtgagagttctagatgaaacaaagaagtaagaataacaattaca 540
Db 362500 GTTAAAGCCAACAGATGTAGAACTCTAGATVGAACAAAAGGTAAACATTAACATTAACA 362441

QY 541 ttaattacttggatgatatcaatgaaagacagcgcttgggaaaaacgtaaatctt 600
Db 362440 TTAATTACTTGGATGATATTACATGAAAAGACAGCGCTTTGGGAAAAACGTAAATCTTT 362381

QY 601 gtagctacagaagtcacaataa 621
Db 362380 GTAGCTACAGAAAGTCAATATA 362360

RESULT 9
US-60-038-697-735
: Sequence 735, Application US/60038697
: GENERAL INFORMATION:
: APPLICANT: Lagace, Robert E.
: APPLICANT: Corley, Neil C.
: APPLICANT: Russo, Frank D.
: TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STAPHYLOCOCCUS AUREUS
: NUMBER OF SEQUENCES: 1027
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
: STREET: 3174 PORTER DRIVE
: CITY: PALO ALTO
: STATE: CALIFORNIA
: COUNTRY: USA
: ZIP: 94304
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: word Perfect 6.1 for Windows/MS-DOS 6.2

```

1 CURRENT APPLICATION DATA:
 2 APPLICATION NUMBER: US/60/038,697
 3 FILING DATE:
 4 CLASSIFICATION:
 5 ATTORNEY/AGENT INFORMATION:
 6 NAME: CERONE, MICHAEL C.
 7 REGISTRATION NUMBER: 39,132
 8 REFERENCE/DOCKET NUMBER: PQ-0001 US
 9 TELECOMMUNICATION INFORMATION:
 10 TELEPHONE: (415) 855-0555
 11 TELEFAX: (415) 845-4166
 12 INFORMATION FOR SEQ ID NO: 735:
 13 SEQUENCE CHARACTERISTICS:
 14 LENGTH: 45487 base pairs
 15 TYPE: nucleic acid
 16 STRANDEDNESS: single
 17 TOPOLOGY: linear
 18 MOLECULE TYPE: CDNA
 19 IMMEDIATE SOURCE:
 20 CLONE: SAC0735
 21
 22 US-60-038-697-735

Query Match	92.1%	Score 572.2	DB 28	Length 45487
Best Local Similarity	99.5%	Pred. No. 2.3e-97		
Matches	574	Conservative	0	Mismatches 3; Indels 0; Gaps 0
OY	1	atgaaataatgacaaatcgatataatgacaatcgctggtggtgacttaaccatgtgca	60	
Db	44911	ATGAAAAATATGACAAATCGATTATATGACAAATCGCTGGTGGTACTTATCTAGAGCA	44970	
OY	61	gcatacttgcttgcacaaaccacatctgcatatctcttcaagataaagataagatgaa	120	
Db	44971	GCATATTTGTTGGCTAAACCCACATATTCGTAATTAATCTTCACGATTAAGATGA	45030	
OY	121	aagatctgacaacatgatataaataatgataaagacagcgagtaagaataaaagacaa	180	
Db	45031	AAGATTGAACATATGATTAATAATGTAAAGAACAGCGAGTAAGATTAATAAGACGAA	45090	
OY	181	gtcaaacctcaaatctcgagaagataaatacgaagtggcaggtcatatgtgaattccagat	240	
Db	45091	GCTAAACCTCAAAATTCGGAAGATTAATTCGAAAGTGGCGGTATATTCAAATTCAGAT	45150	
OY	241	gtgatatataaagacacagatatataccagaccagcaacacctgaacaatlaaatgaagt	300	
Db	45151	GCTGATATTAAAGAACCGACTATATCCAGGACCAACACCTGAAACAATTAATTAAGAGT	45210	
OY	301	gtaaacctctgcagaagaaatgaatcaactagaatgatacaaatatttcaactlccagacac	360	
Db	45211	GTAACCTTTCGAGGAAGAAATGAATCAGTACAGTGAATCAAAATTAATTCATTCGAGCAC	45270	
OY	361	acttccatctgacgcgtccgaactatacaaatcttaaaagcaagccaaaaagtgagt	420	
Db	45271	ACTTTCATTTGACCGCGCCGAACTATCAATTTACAAATCTTAAAGCGACCCAAAAAGGTAT	45330	
OY	421	atggctactcttaaaagtttgatgaataaagcaagtaagtataaaatgacaagataaagaagt	480	
Db	45331	ATGGCTACTTTAAAGTTTGATATATAAACACGCTAAGTATATAATGCAAGTATTAAGAGAT	45390	
OY	481	gttaagcctacagatgttagagttctagatgacacaaaaggtlaaagataaacaattaa	540	
Db	45391	GTTAAGCCTACAGATGTGAAGTCTCGATGTAACAAAAAGGTAAAGATTAACAAATTAACA	45450	
OY	541	ttaattacttctgtagatgattacaatgaaaaaagcaagcg	577	
Db	45451	TTAATTACTTGTGATGATTACATGAAAAAGACAGCGC	45487	
RESULT	10			
	US-60-046-714-734			
	Sequence 734, Application US/60046714			
	GENERAL INFORMATION:			
	APPLICANT: Lagace, Robert E.			

APPLICANT: Corley, Neil C.
APPLICANT: Russo, Frank D.
APPLICANT: Heath, Joe D.
TITLE OF INVENTION: NUCLEOTIDE SEQUENCES OF STAPHYLOCOCCUS AUREUS
TITLE OF INVENTION: GENOME, FRAGMENTS THEREOF, AND USES THEREOF
NUMBER OF SEQUENCES: 1050
CORRESPONDENCE ADDRESS:
ADDRESSEE: INCYTE PHARMACEUTICALS, INC.
STREET: 3174 PORTER DRIVE
CITY: PALO ALTO
STATE: CALIFORNIA
COUNTRY: USA
ZIP: 94304
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Word Perfect 6.1 for Windows/MS-DOS 6.2
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/60/046,714
FILING DATE: HEREWITH
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: CERRONE, MICHAEL C.
REGISTRATION NUMBER: 39,132
REFERENCE/DOCKET NUMBER: PM-0001-1P P
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 855-0555
TELEFAX: (415) 845-4166
INFORMATION FOR SEQ. ID NO: 734:
SEQUENCE CHARACTERISTICS:
LENGTH: 45488 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: cDNA
IMMEDIATE SOURCE:
CLONE: SA01C734
US-60-046-714-734

Query Match 92.1%; Score 572.2; DB 29; Length 45488;
Best Local Similarity 99.5%; Pred. No. 2.3e-97;
Matches 574; Conservative 0; Mismatches 3; Indels 0; Gaps 0;

QY 1 atgaaaaatgacaacatcgatlaaagcaacgcgtggtgactatctagtgca 60
|||||
Db 44912 ATGAAAAATGACAAATCGATTATGACATCGCTGCTGTTGTTACTTATCTTGTGCA 44971
|||||
QY 61 gcatattgttctaaacacacatcgaataattacttcaagataaagatgaa 120
|||||
Db 44972 GCATATTGTTGCTAAACCCATATCGATTATCTTTCACGATTAAGATTAAGATGAA 45031
|||||
QY 121 aagatlaaacatataatgaataatgtaaaagacagcgagtaagtaaaagacgaa 180
|||||
Db 45032 AAGATTACACATATGATTAATAATGTAAGACAGCGAGTAAAGATTAAGATTAAGACGAA 45091
|||||
QY 181 gctaaacctcaaatccgaagaataaactgaagtgagcgatlatgaaatccagat 240
|||||
Db 45092 GCTAAACCTCAAAATTCGAAAGATTAATCGAAATGCGAGCTATATTGAAATTCACAT 45151
|||||
QY 241 gctgaatlaaagaacagatataccagagacagacaacctgtaacaattaaatagagt 300
|||||
Db 45152 GCTGATTTAAAGAACAGTATATCCAGACACCAACACCTGAACATTAATTAAGAGT 45211
|||||
QY 301 gtaagcttgcagaagaataatgaatcgaatgatacaaaatattcaatgacagacac 360
|||||
Db 45212 GTAAAGCTTTGCAGAGAAAGATGATCACTAGATGATCAAAATATTTCATATTGCGAGAC 45271
|||||
QY 361 acttcatlgaacgctcgaacatcaatttacaatctttaaagcagccaaaaagagtagt 420
|||||
Db 45272 ACTTTCATGACCGCTCCGAACATCAATTTTACAATTTTAAGACGCCAAAAAGGTAGT 45331
|||||

QY 421 atggtlactttaagtgtglaaayaaacgcgtaagatlaaaatgacaagatlaagagt 480
|||||
Db 45332 ATGGTACTTTAAAGTTGTAATGAACACGTAAGTAAATGACAAAGTATTAAGACAT 45391
|||||
QY 481 gtaagcttaagaatgtagggttctagatgaacaaaaaggttaagataaaccaatlaa 540
|||||
Db 45392 GTTAAGCCTACAGATGTAAGAGTTCTGATGTAACAAAAAGGTAAAGATTAAGAC 45451
|||||
QY 541 ttaattactgtgagatlaacatgaagaagacagcg 577
|||||
Db 45452 TTAATTACTTGATGATTAACATGAAGAACAGACGCG 45488
|||||

RESULT 11
US-09-620-608-849
Sequence 849, Application US/09620608
GENERAL INFORMATION:
APPLICANT: Chapman, Rowan
APPLICANT: Nuttal, Rachel
TITLE OF INVENTION: POLYNUCLEOTIDES OF STAPHYLOCOCCUS AUREUS
FILE REFERENCE: PA-0016 US
CURRENT APPLICATION NUMBER: US/09/620,608
PRIOR FILING DATE: 2000-07-20
PRIOR APPLICATION NUMBER: 60/144,883
NUMBER OF SEQ. ID NOS: 1999-07-20
SOFTWARE: PERL Program
SEQ. ID NO 849
LENGTH: 551
TYPE: DNA
ORGANISM: Staphylococcus aureus
FEATURE:
NAME/KEY:
OTHER INFORMATION: 813650849
PUBLICATION INFORMATION:
US-09-620-608-849

Query Match 86.0%; Score 534.2; DB 22; Length 551;
Best Local Similarity 99.3%; Pred. No. 1.7e-90;
Matches 547; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

QY 33 cgtctgtgtgttacttaccctagtgagcagcatattgttgcctaacaacacatcgataa 92
|||||
Db 1 cgtctgtgtgttacttaccctagtgagcagcatattgttgcctaacaacacatcgataa 60
|||||
QY 93 ttatcttcagataaagataaagatgaaagattgaaacatataatgtaaaaga 152
|||||
Db 61 ttatcttcagataaagataaagatgaaagattgaaacatataatgtaaaaga 120
|||||
QY 153 acagcgagtaagaataaagaacagcagcaagctaaacctcaaatccgaagaataatcgaa 212
|||||
Db 121 acagcgagtaagaataaagaacagcagcaagctaaacctcaaatccgaagaataatcgaa 180
|||||
QY 213 agtgcaggtctatatgaaattccagatgctgata- ttaaagaaaccagtatataccagac 271
|||||
Db 181 agtgcaggtctatatgaaattccagatgctgata- ttaaagaaaccagtatataccagac 240
|||||
QY 272 cagcaaacacctgaaacatataatagagtgtaagctttgcagaagaagaatgaatcaactag 331
|||||
Db 241 cagcaaacacctgaaacatataatagagtgtaagctttgcagaagaagaatgaatcaactag 300
|||||
QY 332 atgattcaaaatattcaattgacagacacatttcaattgacgctcgaacatacaattta 391
|||||
Db 301 atgattcaaaatattcaattgacagacacatttcaattgacgctcgaacatacaattta 360
|||||
QY 392 caaatcttaagcagcaaaaaagtagtagtgtgtacttaagtgtgtaatgaacac 451
|||||
Db 361 caaatcttaagcagcaaaaaagtagtagtgtgtacttaagtgtgtaatgaacac 420
|||||
QY 452 gtaagataaaatgacaagataaagatgttaaagcttaacagatgttgagagtttagatg 511
|||||

Db 421 gtaagtataaaatgcaagataaagatgttaagcctacacatgtagaagttcttgatg 480
Qy 512 aacaaaagtgataaagataaacaattacattactgtgtatgatactgaatgaaga 571
Db 481 aacaaaagtgataaagataaacaattacattactgtgtatgatactgaatgaaga 540
Qy 572 caggcgcttgg 582
Db 541 caggcgcttgg 551

RESULT 12
US-60-144-883-849
; Sequence 849, Application US/60144883
; GENERAL INFORMATION:
; APPLICANT: Chapman, Rowan
; APPLICANT: Nuttal, Rachel
; APPLICANT: Thornton, Michael
; TITLE OF INVENTION: POLYNUCLEOTIDES OF STAPHYLOCOCCUS AUREUS
; FILE REFERENCE: PA-0016 P
; CURRENT APPLICATION NUMBER: US/60/144,883
; CURRENT FILING DATE: 1999-07-20
; NUMBER OF SEQ ID NOS: 1973
; SOFTWARE: PERL Program
; SEQ ID NO 849
; LENGTH: 551
; TYPE: DNA
; ORGANISM: Staphylococcus aureus
; FEATURE:
; NAME/KEY:
; OTHER INFORMATION: 813650849
; PUBLICATION INFORMATION:
US-60-144-883-849

Query Match 86.0%; Score 534.2; DB 39; Length 551;
Best Local Similarity 99.3%; Pred. No. 1.7e-90;
Matches 547; Conservative 0; Mismatches 3; Indels 1; Gaps 1;

Qy 33 cgcgtgtgtgtactatctacgtatgacagcagatattgttgcgttaaccacatcatgata 92
Db 1 cgcgtgtgtgtactatctacgtatgacagcagatattgttgcgttaaccacatcatgata 60
Qy 93 ttaattcgcga taaagataaagatgaagaatgaacacata tgaataaataatgaaga 152
Db 61 ttaattcgcga taaagataaagatgaagaatgaacacata tgaataaataatgaaga 120
Qy 153 acagcgagtaagataaagaacagcaagctaaacctcaattccgaagaatgaatcgaa 212
Db 121 acagcgagtaagataaagaacagcaagctaaacctcaattccgaagaatgaatcgaa 180
Qy 213 agtgcaggtatatttgaattccagatgtcgata -ttaagaaccagatataccagagc 271
Db 181 agtgcaggtatatttgaattccagatgtcgata -ttaagaaccagatataccagagc 240
Qy 272 cagcaacacctgaacaatttaataagatgtaagcttcgagaagaataatgaatcactag 331
Db 241 cagcaacacctgaacaatttaataagatgtaagcttcgagaagaataatgaatcactag 300
Qy 333 atgatacaaatatttcaattgcagacacaccttcaattccgagccgaaactcaattta 391
Db 301 atgatacaaatatttcaattgcagacacaccttcaattccgagccgaaactcaattta 360
Qy 392 caaattcctaagaacgcaaaaaggtatgattgttactttaaagtgttaataagaacc 451
Db 361 caaattcctaagaacgcaaaaaggtatgattgttactttaaagtgttaataagaacc 420
Qy 452 gtaagtataaaatgcaagataaagatgttaagcctacagatgaaggttcctagatg 511
Db 421 gtaagtataaaatgcaagataaagatgttaagcctacagatgaaggttcctagatg 480
Qy 512 aacaaaagtgataaagataaacaattacattactgtgtatgatactgaatgaaga 571

Db 481 aacaaaagtgataaagataaacaattacattactgtgtatgatactgaatgaaga 540
Qy 572 caggcgcttgg 582
Db 541 caggcgcttgg 551

RESULT 13
US-09-450-969-1132
; Sequence 1132, Application US/09450969
; GENERAL INFORMATION:
; APPLICANT: Lynn Doucette-Stamm et al.
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO STAPHYLOCOCC
; TITLE OF INVENTION: EPIDERMIDIS FOR DIAGNOSTICS AND THERAPEUTICS
; FILE REFERENCE: PAT99-09A
; CURRENT APPLICATION NUMBER: US/09/450,969
; CURRENT FILING DATE: 1999-11-29
; NUMBER OF SEQ ID NOS: 7544
; SEQ ID NO 1132
; LENGTH: 612
; TYPE: DNA
; ORGANISM: S. epidermidis
US-09-450-969-1132

Query Match 54.8%; Score 340.2; DB 18; Length 612;
Best Local Similarity 73.1%; Pred. No. 3.4e-54;
Matches 454; Conservative 0; Mismatches 158; Indels 9; Gaps 1;

Qy 1 atgaaaataatggaacacatcgatgaatgaacacatcgctgtgtgtactatccctagtgaga 60
Db 1 atgaaaataatggaacacatcgatgaatgaacacatcgctgtgtgtactatccctagtgaga 60
Qy 61 gcatattgttgcgttaaccacatcatgataattatcttcacgataaagatga 120
Db 61 attattattctcaagcactatcatgataattatcttcacgataaagatga 120
Qy 121 aagatgacaacatataataaataatgaagaacagcgagtaagataaagaacagca 180
Db 121 aaatgacaacatataataaataatgaagaacagcgagtaagataaagaacagca 171
Qy 181 gctaaacctcaaatctcgaagaataaatacgaagaatgagcagctata ttgaattccagat 240
Db 172 tgaagcgaagaatgatactccgataaataatgagcagctata ttgaattccagat 221
Qy 241 gctgataataaagaacacatata tccagagcagcaacacccgaaacataatgaagat 300
Db 232 gcaacaataaagaacacatata tccagagcagcaacacccgaaacataatgaagat 291
Qy 301 gtaagcttgcagaagaataatgaatcactagatgaatcaaaataattcatttgcagagc 360
Db 292 gtaagcttgcagaagaatgaatcactagatgaatcaaaataattcatttgcagagc 351
Qy 361 acttcatgacgctgcgacataccaatttacaacatcttaagaacagcacaagaatgag 420
Db 352 acgtttacagatcgctgcgacataccaatttacaacatcttaagaacagcacaagaatgag 411
Qy 421 atggttactttaaagttgttaatagaacacgtaagatataaataatgaagaatgaagat 480
Db 412 aaagtgtattttaaactgtgaataccaactagaagaatataaaataactaaataactgtat 471
Qy 481 gttaaagcctcagatgttagagttctgaatgaacaaaagaagataaacaattaaac 540
Db 472 gttaaagcctcagatgttagagttctgaatgaacaaaagaagataaacaattaaac 531
Qy 541 ttaattactgtgtatgatacaatgaagaacagcgcttctgggaaaacgtaaatcttc 600
Db 532 ttaattactgtgtatgatacaatgaagaacagcgcttctgggaaaacgtaaatcttc 591
Qy 601 gtaagcacaagaagtaacata 621
Db 592 atagctacacaaatlaactaa 612


```
OY 1 atgaaacatgacaaatcgatlaaagcaatcgctggtgtgtactatccctagtgca 60
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 1995 ATGAAGCAGATGATGATTAATCACTTAATAGGCGTATGTTATCATTTTAGCT 2054
OY 61 gcaatttggttgcaaacccatcgcataaattcttcacgataaagatgaa 120
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2055 ATTTATTTATTCCTCAACCGCATATTCGATTAATTAATCTACATGAANAAGATPACGATCAT 2114
OY 121 aagattgacaaatcgtatcaaaatgtaaaagacagcgagtaaaagataaaagcagca 180
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2115 AAAATTGAAATTTATGATTAATAAGAAAAGACAGACAAAGACATCTTA-----AA 2165
OY 181 gctaaacctcaaatccgaaagataaatcgaagtgcaggtatattgaaattccagat 240
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2166 TCGACGCCAAAGATACCTTCGATTAATCTAAATGCGTGTATATAGAACTCCAGAC 2225
OY 241 gctgatattaagaaccgattatccagagaccgacaacacctgaacaatlaaataagat 300
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2226 GCACAAATPAAAGAACAGATATACCTGCTCCAGCACACGAAACAACTCAATAGAGGT 2285
OY 301 gtaagcttgcagaagaatcgaatcactagaatgataaataattcaattgcagagacac 360
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2286 GTAAGTTTGGCAGAGGTGATGATCTCTATATCAACAGAAATATTCAATTGCTGTGCTAT 2345
OY 361 acttcaattgacccgctcgaactatacaatttacaactctaaagcgacaaagagtagt 420
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2346 ACGTTTACAGATCGTTCGACACTATCAATTTCAAAATTTAAATTCAGCCAAATCGGTACT 2405
OY 421 atggtgtaacttaaaagtgtgtaataaacagtaagtaataaataagtaagtaagat 480
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2406 AAAGTGTATTTTAAACGTGAATCAAACTAGAAAGTATTAANATACTPAAATACGTGAT 2465
OY 481 gtaagcctacagatgtaaggattctagaatgacaaaagtaaaagataaacaattaca 540
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2466 GTTAGCCTACAGAGTAAAGTCTTAGAGCAACATCCTAATTAAGAAAATCAATTANCA 2525
OY 541 ttaattacttgatgattacaatgaaaagacagcggttgggaaa-aacgtaaaatcct 599
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2526 TTAATTACTTGGATGACTACAAAGAACGCGGTGTTGGAAACCAAGGAAAATATT 2585
OY 600 tgaactacagaagtaaataa 621
    ||||| ||||| ||||| ||||| ||||| ||||| ||||| ||||| |||||
Db 2586 CATAGCTACACAAATTAACCTAA 2607
```

Search completed: March 14, 2001, 15:06:48
Job time: 4688 sec

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 07:35:20 ; Search time 75.42 Seconds
(without alignments)
3093.165 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
Sequence: 1 atgaaaaaatgacaaatcg.....tagctacagaagcaataaa 621

Scoring table: IDENTITY_NUC
Gapox 10.0, Gapext 1.0

Searched: 480022 seqs, 187831343 residues

Total number of hits satisfying chosen parameters: 960044

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

N.Geneseq_36:*

- 1: /SID6/gcgdata/geneseq/geneseqn/NA1980.DAT:*
- 2: /SID6/gcgdata/geneseq/geneseqn/NA1981.DAT:*
- 3: /SID6/gcgdata/geneseq/geneseqn/NA1982.DAT:*
- 4: /SID6/gcgdata/geneseq/geneseqn/NA1983.DAT:*
- 5: /SID6/gcgdata/geneseq/geneseqn/NA1984.DAT:*
- 6: /SID6/gcgdata/geneseq/geneseqn/NA1985.DAT:*
- 7: /SID6/gcgdata/geneseq/geneseqn/NA1986.DAT:*
- 8: /SID6/gcgdata/geneseq/geneseqn/NA1987.DAT:*
- 9: /SID6/gcgdata/geneseq/geneseqn/NA1988.DAT:*
- 10: /SID6/gcgdata/geneseq/geneseqn/NA1989.DAT:*
- 11: /SID6/gcgdata/geneseq/geneseqn/NA1990.DAT:*
- 12: /SID6/gcgdata/geneseq/geneseqn/NA1991.DAT:*
- 13: /SID6/gcgdata/geneseq/geneseqn/NA1992.DAT:*
- 14: /SID6/gcgdata/geneseq/geneseqn/NA1993.DAT:*
- 15: /SID6/gcgdata/geneseq/geneseqn/NA1994.DAT:*
- 16: /SID6/gcgdata/geneseq/geneseqn/NA1995.DAT:*
- 17: /SID6/gcgdata/geneseq/geneseqn/NA1996.DAT:*
- 18: /SID6/gcgdata/geneseq/geneseqn/NA1997.DAT:*
- 19: /SID6/gcgdata/geneseq/geneseqn/NA1998.DAT:*
- 20: /SID6/gcgdata/geneseq/geneseqn/NA1999.DAT:*
- 21: /SID6/gcgdata/geneseq/geneseqn/NA2000.DAT:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	611	98.4	3733	18	Staphylococcus aur
2	67.2	10.8	3399	17	Staphylococcus aur
3	66.8	10.8	1686	16	Staphylococcus aur
4	66.4	10.7	4590	7	DNA encoding leuco
5	59.8	9.6	6644	20	Sequence encoding
6	59.8	9.6	7372	20	Base sequence of t
7	59.8	9.6	7797	20	Base sequence of t
8	59.8	9.6	7996	20	Base sequence of t
9	56.6	9.1	3095	11	Sequence encoding
10	56.6	9.1	5361	18	P. falciparum live
11	56.6	9.1	6152	18	P. falciparum live
12	52.8	8.5	9789	17	cDNA encoding Plas

13	51.4	8.3	910715	20	X20248
14	51.2	8.2	4940	19	V35363
15	51	8.2	31096	18	V74370
16	50.8	8.1	839	10	N90127
17	50.6	8.1	9636	15	067190
18	50.2	8.1	1612	7	N60392
19	49.6	8.0	5760	6	N50530
20	49	7.9	389	5	N40273
21	48.8	7.9	1956	18	T67161
22	48.4	7.8	4248	21	Z56910
23	48.4	7.8	6361	21	Z91806
24	48.2	7.8	1368	15	Q70181
25	48.2	7.8	1425	15	Q70182
26	48.2	7.8	1524	15	Q70183
27	48.2	7.8	3294	15	Q70179
28	48	7.7	846	19	Z66473
29	48	7.7	3132	19	V52243
30	47.8	7.7	3300	21	Z56939
31	46.8	7.5	58407	19	V21210
32	46.4	7.5	1035	21	A01936
33	46.4	7.5	1300	18	T84232
34	46.4	7.5	1300	19	V53553
35	46.2	7.4	2223	16	O80908
36	46.2	7.4	2685	21	Z43781
37	46.2	7.4	19124	18	T72882
38	46.2	7.4	19124	21	Z98287
39	45.8	7.4	1902	21	Z91851
40	45.8	7.4	9541	19	V52265
41	45.4	7.3	1590	20	Z31963
42	45.4	7.3	1680	20	Z31961
43	45.2	7.2	2405	20	Z31958
44	45	7.2	1410	20	X61755
45	45	7.2	12730	19	V44233

ALIGNMENTS

RESULT 1
V74849/c
ID V74849 standard; DNA: 3733 BP.
XX
AC V74849:
XX
DT 16-MAR-1999 (first entry)
XX
DE Staphylococcus aureus contig SEQ ID #538.
XX
XX Computer readable medium; vaccine; S. aureus infection; immunodetection;
KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
KW skin infection; surgical wound infection; scalded skin syndrome;
KW toxic shock syndrome; ds.
XX
OS Staphylococcus aureus.
XX
FH Key Location/Qualifiers
FT misc_feature 721..780
FT FT /tag= a
FT FT /note= "these bases represent a line of missing text in
the sequence listing to maintain the nucleotide numbering
given in the specification for this DNA sequence"
FT misc_feature 2521..2580
FT FT /tag= b
FT FT /note= "these bases represent a line of missing text in
the sequence listing in the specification. They
are included to maintain the nucleotide numbering
given in the specification for this DNA sequence"
XX
XX EP786519-A2.
XX
XX 30-JUL-1997.
XX

PE 07-JAN-1997; 97EP-0100117.
 XX 05-JAN-1996; 96US-0009861.
 XX
 PA (HUMA-) HUMAN GENOME SCI INC.
 XX
 PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 PI Rosen CA;
 XX
 DR WPI; 1997-374922/35.
 XX
 PT Polynucleotide(s) and proteins derived from *Staphylococcus aureus*
 PT stored on computer readable medium and used in the production of
 PT anti-*S. aureus* vaccines
 XX
 XX
 Claim 1; Page 1472-1474; 3271pp; English.
 OS

This sequence represents one of 5191 *Staphylococcus aureus* DNA sequences of the invention. The DNA sequences are recorded on a computer readable medium, preferably selected from a floppy or hard disk, random access memory (RAM), read-only memory (ROM) or CD-ROM. Homology searches using the S.aureus DNA sequences allows putative functions to be assigned so that protein-encoding or regulatory regions of commercial, therapeutic or industrial importance can be obtained. Specifically, sequences which are likely to encode antigens have been identified and these polypeptides can be used in a vaccine composition against S.aureus infection. The polypeptides can also be used in a kit for the immunodetection of S.aureus in a sample. S.aureus is implicated in numerous human diseases, including cellulitis, eyelid infections, food poisoning, osteomyelitis, skin and surgical wound infections, scalded skin syndrome, toxic shock syndrome, etc. Organisms transformed with the DNA sequences can be used for recombinant production of the polypeptides. The new DNA sequences (and their fragments) are useful as primers or probes for isolating homologues of any of the S.aureus DNA sequences contained on the computer readable medium.

Sequence 7733 BP; 1084 A; 718 C; 543 G; 1266 T; 122 other;

Query Match	98.4%	Score 611	DB 18	Length 3733
Best Local Similarity	100.0%	Pred. NO.	1.2e-118	
Matches 611	Conservative 0	Mismatches 0	Indels 0	Gaps 0

OY	11	ggaaaatgatttaatgacaatcgccggtgtggtactatccctggtggcgcatattgt	70
Db	720	GGCAAAATCGATTATATACAAATCGCTGTGTGTGACTTATCTAGTGGCAGCAATTTGT	661
OY	71	ttgttaaccacatcatgataattatcttcacga taaaga taaagatgaaagattgaac	130
Db	660	TTGTGAATCCCATATGTGATATATTATCTTACGTTAAAGTAAAGTAAAGATTTCAC	601
OY	131	aata tga taaaaatg taaaaagcaaggcgag taaaga taaaaagcagcaagcttaaacctc	190
Db	600	AATGTGATAAAAATGTAAAGAACAGCGCGAGTAAAGTAAAAAGCAGCAAGCTAAACCTC	541
OY	191	aaatctcgaaagataaa tgcgaagttggcaggtcatatltgaatccagatgcgtgata tta	250
Db	540	AAATTCGGAAGATTAATTCGAAAGTGGCAGCGCTATATTGAAATTCGAGATCGATATT	481
OY	251	aagaacacagatataccagagccagcaacacctbaaacaataaataagagtgtaagctttg	310
Db	480	AAGAACACAGTATATCCAGGACCGCAACACCTTAACATTTAAATVAGGGTGTAAAGCTTTG	421
OY	311	cagaagaana tgaatcaatcagatgataccaataat tccaatctgcaagcaaaccttcatg	370
Db	420	CAGAAGAATAATGATCTCATGATGATATAAATTTTCAATTTCGACGACACACTTTCAT	361
OY	371	accgtccgaactlcaatlttacaacatctlaaagcagccaaaagaagtagtatgttact	430
Db	360	ACGGTCGCACTATCAATTTACAAATCTTAAGCAGCCAAAAGAGTAGTATGCTACT	301
OY	431	ttaaagttggtaatgaaacacg tlaagla taaaatgacaag taaagaatg ttaagacta	490

Db	300	TTAAGTTGGTAATGAAACACGTAAGTATATAAAGACAGTATTAAGAGATTAAAGCCTA	241
Qy	491	cagaatgttagaggtctcagatgtaacaaaaaggttaagaataacaattacaataact	550
Db	240	CAGATGTAGGAGTCTCTGATGAACAAAGGTAAGATAACATTAATTAATCTT	181
Qy	551	gtgatgtattacaatgaaaaagacagcgctttgggaaaaagtaaatctttgtagctacag	610
Db	180	GTCGATGATTACAAATGAACAAAGACAGCGCTTGGGAAAAACCTAAATCTTGTAGCTACAG	121
Qy	611	aaatcaaataa	621
Db	120	AAGTCAAAATA	110

RESULT	2	
ID	T05868	
XX	T05868 standard; DNA: 3399 BP.	
XX		
AC	T05868;	
XX		
DT	14-AUG-1996 (first entry)	
XX		
DE	Chicken leucocytozoan DNA, encoding immunogenic protein for vaccines	
XX		
KM	Chicken leucocytozoan; immunogen; recombinant vaccine; protection;	
KW	immunisation; vaccination; ss.	
XX		
OS	Chicken leucocytozoan.	
XX		
FH	Key	Location/Qualifiers
FT	CDS	1..3399
FT		/*tag= a
FT	misc_feature	1150..3318
FT		/*tag= b
FT		/note= "fragment referred to in the claims, for use as insert in a recombinant vaccine against chicken leucocytozoan disease"
FT		

PN	JP07284392-A.
XX	
PD	31-OCT-1995.
XX	
PF	19-APR-1994; 94JP-0080643.
XX	
PR	19-APR-1994; 94JP-0080643.
XX	
PA	(DOBU-) DOBUTSUO SEIBUTSUGAKUTEKI SEIZAI KYOKAI.
PA	(KITA) KITASATO KENKYUSHO SH.
XX	
DR	WPI; 1996-006311/01.
DR	P-PSDB; R97866.
XX	
PT	Chicken leucocytozoan immunogenic protein - used in a recombinant vaccine against chicken leucocytozoan disease
XX	
XX	Claim 6; Page 6-9; 35pp; Japanese.
PS	
XX	
CC	T05868 encodes a chicken leucocytozoan immunogenic protein, this DNA or a fragment of it can be used in a recombinant vaccine to immunise against chicken leucocytozoan disease. The DNA is used in a vector and operatively linked to an expression regulatory sequence as in standard practice.
CC	
CC	
XX	
XX	Sequence 3399 BP; 1577 A; 508 C; 798 G; 516 T; 0 other;

Query Match	10.8%	Score 67.2	DB 17	Length 3399
Best Local Similarity	45.6%	Pred. NO. 1e-05		
Matches 237	Conservative	0	Mismatches 283	Indels 0
101	acgataaagataaagatgtaaaagattgaacatatgtataaaaaagttaaaagacagcgca	160		

Db 2069 aagaaagaagaataacacatgaagaataagaagaagaagcatgaagaatc 2128
 Qy 161 gtaagaataaagaagcagcagcctaaacccaatccgaagaataatcgaaagtcag 220
 Db 2129 atgaagaagaagaagaagaataacacatgaagaagaagaagaagaagcctgaagaag 2188
 Qy 221 gctatattgaattccagatctgatataaagaacagtatataccagaccagcaaac 280
 Db 2189 taatcatgaagaagaagaagaagaataacacatgaagaagaagaagaagaataac 2248
 Qy 281 ctgaacaattaatagaagtgtaagcttgcaagaagaagaataatcactagatgtaaa 340
 Db 2249 atgaagaagaagaagaagaataacacatgaagaagaagaagaagaagaataatcagag 2308
 Qy 341 atattcattcagcagcagcacttccatgacgcctccgacatacattacaaatccta 400
 Db 2309 aagaaagaagaagaagaataacacatgaagaagaagaagaagaataacacatgaagaaga 2368
 Qy 401 aagcagcgaagaagaagtgatgctgctgaatgaagtgtaagaacagtaagata 460
 Db 2369 aagaagaagtaacacatgaagaagaagaagaagaagaataacacatgaagaagaagaag 2428
 Qy 461 aatgaacaagataagaagatgttaagcctacagatgtagagcttctagaatgaacaaag 520
 Db 2429 aagtaacacatgaagaagaagaagaagaataacacatgaagaagaagaagaagaataac 2488
 Qy 521 gtaagaataaacaattacattacttgatgcatgaatgaagaagaagcgtt 580
 Db 2489 atgaagaagaagaagaagaagaataacacatgaagaagaagaagaagaataacacatgaagaag 2548
 Qy 561 gggaagaacgtaaaatcttctgagctacagaagtcnaata 620
 Db 2549 aagaagaagtaacacatgaagaagaagaagaagaataacata 2588
 RESULT 3
 087587
 ID 087587 standard; DNA: 1686 BP.
 AC 087587;
 XX
 Dt 19-DEC-1995 (first entry)
 XX
 DE DNA encoding Leucocytozoan protozoa structural protein epitope.
 XX
 KW Leucocytozoan protozoa: structural protein; epitope; vaccine; fowl;
 XX
 OS Leucocytozoan protozoa sp.
 XX
 JN JP0708995-A.
 XX
 PD 04-APR-1995.
 XX
 PF 10-SEP-1993; 93JP-0226078.
 XX
 PR 10-SEP-1993; 93JP-0226078.
 XX
 PA (DOBU-) DOBUTSUYO SEIBUTSUKUTTEKI SEIZAI KYOKAI.
 XX
 PA (N1SS-) N1SSEIKEN KK.
 XX
 DR WPI; 1995-167252/22.
 XX
 P-PSDB; R70491.
 XX
 PT Immune inducing polypeptide against Leucocytozoan protozoa - useful
 XX
 PT in production of vaccines for treatment of Leucocytozanosis in
 XX
 PS fowl.
 XX
 PS Claim 1; Page 12-14; 20pp; Japanese.
 CC 087587-89 encode polypeptides having a whole or partial epitope of a
 CC structural protein of Leucocytozoan protozoa (see R70491-93). The
 CC polypeptides and DNA encoding them are useful in the production of

CC vaccines for the treatment of Leucocytozanosis of fowl.
 XX
 SQ Sequence 1686 BP; 915 A; 129 C; 382 G; 260 T; 0 other;
 Query Match 10.8%; Score 66.8; DB 16; Length 1686;
 Best Local Similarity 45.2%; Pred. No. 1.1e-05;
 Matches 245; Conservative 0; Mismatches 297; Indels 0; Gaps 0;
 Qy 77 aaccacatcgaataattcttcacgaataaagaatgaagaatgaagaatgaagaatga 136
 Db 188 aagaacaagatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 247
 Qy 137 ataaaatgttaaagaagcagcgcagtaagaatlaaagaagcagcagcctaaacccaatcc 196
 Db 248 aag 307
 Qy 197 cgaagaataatcgaagaagtcgacgctataatgaattccagaatgctgatatgaagaac 256
 Db 308 aag 367
 Qy 257 cagtatatccagacagcagaacacacccgaacaatttaagaagtgtaagccttgcaagaag 316
 Db 368 aag 427
 Qy 317 aaatgaatcactagaatgaatgaatgaatgaatgaatgaatgaatgaatgaatgaatga 376
 Db 428 aacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 487
 Qy 377 cgaactatcaattacaaatcttaagcagccaaagaagtagtatgctgactttaag 436
 Db 488 aacaaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 547
 Qy 437 ttggtatgaacacacgtaagataaataatgaagaatgaagaatgttaagcctacagatg 496
 Db 548 aagaacaagaagaagaatgatctgcgaagaagaagaagaagaagaagaagaagaagaagaag 607
 Qy 497 taggaattctagaatgaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 556
 Db 608 aag 667
 Qy 557 attacaatgaagaagcagcgttgggaagaagaagaagaagaagaagaagaagaagaagaag 616
 Db 668 tagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaagaag 727
 Qy 617 aa 618
 Db 728 aa 729
 RESULT 4
 N60472
 ID N60472 standard; DNA: 4590 BP.
 AC N60472;
 XX
 Dt 24-AUG-1991 (first entry)
 XX
 DE Sequence encoding the ring-infected Erythrocyte Surface Antigen
 XX
 DE (RESA).
 XX
 KW Malaria vaccine; antigen; epitope; ss.
 XX
 OS Plasmodium falciparum.
 XX
 FH Key Location/Qualifiers
 XX
 FT exon 801..995
 FT exon /*tag= a
 FT exon 1199..4225
 FT exon /*tag= b
 XX
 PN W08601802-A.

CC plasmid prx-Bcl 2-i-hcd 25, which contains the human Bcl-2 gene, and
CC is used in an example from the present invention.

Sequence 7996 BP; 2463 A; 2015 C; 1829 G; 1689 T; 0 other;

Query Match 9.6%; Score 59.8; DB 20; Length 7996;
Best Local Similarity 44.3%; Pred. No. 0.00039;
Matches 244; Conservative 0; Mismatches 307; Indels 0; Gaps 0;

```
QY 71 ttgcataaccacatcgtatcgtatcctccagataagataagataagattgaac 130
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5087 ttgcataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5146
QY 131 aatattgataaaatgtaaaaagacagcgagtaagataaaagcagcaagctaaaccc 190
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5147 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5206
QY 191 aaattccgaagataaattcgaaagtgccaggtatattgaattccagatgctgatat 250
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5207 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5266
QY 251 aagaaaccgatatatccagacagcaccccgaaacattaaatagagtgtaagcttg 310
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5267 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5326
QY 311 cagaagaataatgaatcactagatgataatattcaattcgagcagcacttcattcg 370
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5327 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5386
QY 371 accgtccgaactatacaattacaalcttaagcagccaaaaaggtagtagtgtagt 430
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5387 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5446
QY 431 ttaaggttgtaagtgaacacgtaagataaaatgcaagataagagtgtaagctta 490
   ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5447 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5506
QY 491 cagatgtagaggtcttagatgaacaaaagtaagataaacattacattactt 550
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5507 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5566
QY 551 gtgattgatacaatgaagaacagcgcttggtgaaacgtaaaatctttagctacg 610
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5567 aaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaaataaaaaa 5626
QY 611 aagtcataaa 621
   || ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 5627 aaaaaaataaaaaa 5637
```

RESULT 9
003875
ID 003875 standard; DNA; 3095 BP.
XX
XX Q03875;
XX
XX 24-AUG-1990 (first entry)
XX
XX Sequence encoding carboxylic terminal part of native GLURP.
XX
XX Plasmodium falciparum; antigen; malaria; vaccine; GLURP:ss
XX
XX Plasmodium falciparum.
XX
XX
XX Key Location/Qualifiers
XX FT J..2352
XX FT /*tag=a
XX FT /product=GLURP
XX PN W09022811-A.
XX PD 22-MAR-1990.

```
XX PF 18-SEP-1989; 89WO-0000218.  
XX PR 03-MAR-1989; 89US-0218885.  
XX PR 03-MAR-1989; 89DK-0005191.  
XX  
XX (STAT-) STATENS SERUMINST.  
XX  
XX Driegel M, Borre M, Jepsen S, Vuust J, Rieneck K, Wind A;  
XX Jakobsen PH;  
XX  
XX WPI: 1990-115998/15.  
XX P-PSDB: R05804.  
XX  
XX POLYPEPTIDE(S) derived from Plasmidium falciparum antigen - used in  
XX vaccines and in production of antibodies, for diagnosis and  
XX therapy of malaria.  
XX  
XX Disclosure: Fig 7; 108pp; English.  
XX  
XX An open reading frame of 2349 bps extends from the 5' terminal end of the  
XX insert to a "TAA" stop codon. It is longest ORF found in the sequence.  
XX Sequence displays some of the characteristics of other malaria nucleic  
XX acid sequences : tandemly repeated motifs, high AT content, and a high  
XX content of codons for glutamate. Three major repetitive sequences are:  
XX one motif from bp 34 to bp 156 is repeated from bp 289 to bp 411; another  
XX motif from bp 477 to bp 521 is repeated from bp 1174 to bp 1233 is  
XX repeated from bp 567 to bp 611; a third motif from bp 1174 to bp 1233 is  
XX repeated tandemly 11 times. This last repetitive region consists of 360bp  
XX repeats differing only in 3 bases GAT coding for aspartate. This region  
XX is flanked to the 5' terminal od a degenerated 60 bp repeat. GC content  
XX of the coding part of the insert is on average 30%, and of the non-coding  
XX 3' terminal 11%  
XX  
SQ Sequence 3095 BP; 1443 A; 300 C; 491 G; 861 T; 0 other;
```

Query Match 9.1%; Score 56.6; DB 11; Length 3095;
Best Local Similarity 44.0%; Pred. No. 0.0016;
Matches 239; Conservative 0; Mismatches 304; Indels 0; Gaps 0;

```
QY 72 tgcataaccacatcgtatcgtatcctccagataagataagataagattgaac 131
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1260 tgaatagtagaggttggaagaattctaccagaagatgataaaatgaaaaggtcac 1319
QY 132 atattgataaaatgtaaagaacagcgagtaagataaaagcagcaagcttaacct 191
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1320 tgaatagtagaggttggaagaattctaccagaagataaaatgaaaaggtcacatga 1379
QY 192 aattccgaagaataatcgagaagtggtgaggtctattgaattccagatgctgattaa 251
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1380 aatagtagaggttggaagaattctaccagaagataaaatgaaaaggttgaacattga 1439
QY 252 agaacagatataccaggacgaacacactgacaaatlaaataagagtgtagctttgc 311
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1440 agtagaagttggaagaattctaccagaagataaaatgaaaaggttcaaatgaatagt 1499
QY 312 agagaagaatgtaacactagatgatacaaaatctcaattgcagcgacacattcatga 371
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1500 agaggttggaagaattctaccagaagataaaatgaaaaggttcaaatgaatga 1559
QY 372 ccgtccgaactatcaattcaaatcttaagagcagcaaaaaggttagtagtgtagt 431
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1560 agttgagaagaattctaccagaagataaaatgaaaaggttcaaatgaatgaatga 1619
QY 432 taaagttgtaacgaaacagtaagataaaatgaaacagataagagtttaagcttac 491
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1620 tgaagaagaattctaccagaagataaaatgaaaaggttcaaatgaatgaatga 1679
QY 492 agatgtagaggttctagatgaacaaaaggtaagataaaacattacattacttg 551
   || | ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1680 tgaagaagaattctaccagaagataaaatgaaaaggttcaaatgaatgaatga 1739
```



```

OY 486 gccctacagatgtagagctcttagatgaacacaaagtgtaagaataacattacattaat 545
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345650 ccccttcaatgaagaatataacacatctaaagaaatctttaaagaacgaanaata 345709
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 546 tactctgagtgatgataatgaataa 568
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 345710 ttaigtctaatgaataaaaaa 345732
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 14
V35363
ID V35363 standard; DNA: 4940 BP.
XX
XX V35363;
XX
XX 23-SEP-1998 (first entry)
XX
DE P. falciparum gp190 DNA.
XX
XX gp190; malaria; MSP-1; merozoite surface protein; stability; vaccine;
XX monoclinal antibody; passive immunisation; parasite; ss.
XX
XX Plasmodium falciparum.
XX
XX Key Location/Qualifiers
XX CDS 10..4929
XX /*tag= a
XX /*product= gp190
XX
XX W09814583-A2.
XX
XX 09-APR-1998.
XX
XX 02-OCT-1997; 97WO-EP05441.
XX
XX 02-OCT-1996; 96DE-4040817.
XX
XX (BUJA/) BUJARD H.
XX
XX Bujard H, Pan W, Tolle R;
XX
XX WPI: 1998-240088/21.
XX P-PSDB; W54145.
XX
XX Recombinant production of complete gp190/MSP-1 Plasmodium surface
XX protein - useful in anti-malaria vaccines, also stabilising genes by
XX reducing their AT content
XX
XX Example 1; Fig 3c; 48pp; German.
XX
XX This sequence encodes the Plasmodium falciparum gp190/MSP-1 (merozoite
XX surface protein). This gene is used in a method for stabilising the
XX gene sequences by reducing the AT content. Such products are useful in
XX vaccines against malaria or for producing monoclonal antibodies (for
XX passive immunisation). The complete gp190 protein can now be produced
XX outside the parasite and has, at least over extended regions, the native
XX pattern of folding. Larger amounts of the protein can be produced
XX recombinantly than would be possible using the parasites as source.
XX
XX Sequence 4940 BP; 2196 A; 597 C; 687 G; 1460 T; 0 other;

Query Match 8.2%; Score 51.2; DB 19; Length 4940;
Best Local Similarity 45.9%; Pred. No. 0.023;
Matches 213; Conservative 0; Mismatches 248; Indels 3; Gaps 1;
OY 104 ataagataaagatgaagaatgaaacaataatgataaaaatgtaagaacagcgagta 163
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1358 aaaaatcctcaataatgaataaagaaaaataaataagaataaaaaaaatgtaatcg 1417
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 164 aagataaaagacgaagctaaacctcaattccgaaataaataatgaaagtcgagct 223
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1418 ataataaatcttaagaaagatctaaagctttaaataatgataaacaagaatagtaa 1477

```

```

OY 224 atattgaattccagatgctgatattaaagaaccaglatatccaggaccagcaacactg 283
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1478 aattacttaattgaatttatagtagcaaatltaataatataatagatttaacttaattcg 1537
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 284 aacaatataatagagtgtaagcttgcagaagaataatgatactcagatgataaata 343
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1538 aaaa--aatgatggtlaaaagataatcatalaagttgagaacattacacacataata 1594
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 344 ttccaattgcagacacacttcatctgaccgctccgacataatcttaacatcttaaga 403
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1595 ctttgcacctcctgaataattcacaataatctgaaaaagttacaacaaagctcttaaat 1654
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 404 cagccaaaaggtagatggtgtactttaaagttgttaatgaacacgtaagtataaa 463
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1655 atatggaagatattcttctaagaatagatagtagtgaataaagaatacaaatatataaa 1714
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 464 tgacaaglatagaagatgttaagcctacagatgtagagttctagatgaacaaaagta 533
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1715 atttaataagcaaaatagaanaatgagattgaaacattagtgtaaaaaatataaaagatg 1774
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
OY 524 aagataaacaataacattacttctgtagatgattacaaatga 567
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 1775 aagaacagcttttgaaaaaaaattactaaagaacgaanaataa 1818
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

RESULT 15
V74370
ID V74370 standard; DNA: 31096 BP.
XX
XX V74370;
XX
XX 16-MAR-1999 (first entry)
XX
XX Staphylococcus aureus contig SEQ ID #59.
XX
XX Computer readable medium; vaccine; S.aureus infection; immunodetection;
XX cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
XX skin infection; surgical wound infection; scalded skin syndrome;
XX toxic shock syndrome; ds.
XX
XX Staphylococcus aureus.
XX
XX Key Location/Qualifiers
XX FH 1201..1260
XX FT /*tag= a
XX FT /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX
XX misc_feature 3001..3060
XX FT /*tag= b
XX FT /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX
XX misc_feature 4801..4860
XX FT /*tag= c
XX FT /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX
XX misc_feature 6601..6660
XX FT /*tag= d
XX FT /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX
XX misc_feature 8401..8460
XX FT /*tag= e
XX FT /note= "these bases represent a line of missing text in
XX the sequence listing in the specification. They
XX are included to maintain the nucleotide numbering
XX given in the specification for this DNA sequence"
XX
XX FT
XX FT
XX FT

```

FT			given in the specification for this DNA sequence
FT	misc_feature	10201..10260	
FT	/tag- f		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	12001..12060	
FT	/tag- g		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	13801..13860	
FT	/tag- h		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	15601..15660	
FT	/tag- i		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	17401..17460	
FT	/tag- j		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	19201..19260	
FT	/tag- k		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	21001..21060	
FT	/tag- l		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	22801..22860	
FT	/tag- m		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	24601..24660	
FT	/tag- n		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	26401..26460	
FT	/tag- o		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	28201..28260	
FT	/tag- p		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	
FT	misc_feature	30001..30060	
FT	/tag- q		
FT	/note-	"these bases represent a line of missing text in the sequence listing in the specification. They are included to maintain the nucleotide numbering given in the specification for this DNA sequence"	

[illegible]

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:21:38 ; Search time 11.91 Seconds
(without alignments)
310.592 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKRWTKMLMTIAGVILVA.....YNEKGVWEKRIEVAEVRK 206

Scoring table:
BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 174772 seqs, 17957048 residues

Total number of hits satisfying chosen parameters: 174772

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :
1: /cgn2_6/ptodata/2/1aa/5A.COMB.pep:*
2: /cgn2_6/ptodata/2/1aa/5B.COMB.pep:*
3: /cgn2_6/ptodata/2/1aa/6.COMB.pep:*
4: /cgn2_6/ptodata/2/1aa/PCRTUS.COMB.pep:*
5: /cgn2_6/ptodata/2/1aa/Backfiles1.pep:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	137	12.7	223	3	US-08-961-083-162
2	89.5	8.3	411	2	US-08-741-134-6
3	88.5	8.2	289	3	US-08-961-083-20
4	88.5	8.2	309	2	US-08-715-131-2
5	82.5	7.7	783	2	US-08-922-837-2
6	82	7.6	427	3	US-09-196-857-2
7	82	7.6	558	1	US-08-285-440-6
8	82	7.6	558	1	US-08-630-349-6
9	80.5	7.5	759	1	US-08-676-967-1
10	80.5	7.5	759	1	US-08-676-974-1
11	80.5	7.5	759	2	US-09-098-487-1
12	80	7.4	190	3	US-08-881-094-20
13	80	7.4	257	2	US-08-715-204-6
14	80	7.4	257	3	US-09-162-597-6
15	80	7.4	273	1	US-08-137-175A-5
16	80	7.4	273	3	US-08-479-017-5
17	79.5	7.4	339	1	US-08-431-080-18
18	79.5	7.4	339	2	US-08-938-534-18
19	79.5	7.4	351	2	US-08-929-417-2
20	79.5	7.4	765	2	US-08-663-112-2
21	79	7.3	153	3	US-08-677-778B-1
22	79	7.3	338	2	US-08-441-857-10
23	79	7.3	338	3	US-08-193-159-10
24	79	7.3	648	1	US-08-451-715A-4
25	78	7.2	426	3	US-08-961-083-48
26	78	7.2	581	3	US-08-961-083-132
27	77	7.2	278	3	US-08-961-083-94
28	77	7.2	451	1	US-08-435-454-4

29	77	7.2	451	1	US-08-652-972A-4	Sequence 4, Appli
30	77	7.2	451	2	US-08-919-145-6	Sequence 6, Appli
31	77	7.2	451	3	US-08-870-126-4	Sequence 4, Appli
32	77	7.2	451	4	US-09-344-889-6	Sequence 6, Appli
33	77	7.2	451	4	PCR-US96-06231A-4	Sequence 4, Appli
34	77	7.2	571	3	US-08-961-083-4	Sequence 4, Appli
35	77	7.2	714	2	US-08-990-114-3	Sequence 3, Appli
36	76.5	7.1	331	2	US-08-828-242-3	Sequence 3, Appli
37	76.5	7.1	331	2	US-08-910-927B-5	Sequence 5, Appli
38	76.5	7.1	1964	2	US-08-790-912-3	Sequence 3, Appli
39	76.5	7.1	2052	2	US-08-790-912-2	Sequence 2, Appli
40	76	7.1	1664	2	US-08-642-846-2	Sequence 2, Appli
41	75.5	7.0	1674	2	US-08-968-542C-12	Sequence 12, Appli
42	75	7.0	593	1	US-08-296-362-2	Sequence 2, Appli
43	74.5	6.9	274	3	US-08-937-271-6	Sequence 6, Appli
44	74.5	6.9	294	1	US-08-137-175A-9	Sequence 9, Appli
45	74.5	6.9	294	3	US-08-479-017-9	Sequence 9, Appli

ALIGNMENTS

RESULT 1
US-08-961-083-162
Sequence 162, Application US/08961083
Patent No. 6159469
GENERAL INFORMATION:
APPLICANT: Choi et. al.
TITLE OF INVENTION: Streptococcus pneumoniae Antigenes and Vaccines
NUMBER OF SEQUENCES: 452
CORRESPONDENCE ADDRESS:
ADDRESSEE: Human Genome Sciences, Inc.
STREET: 9410 Key West Avenue
CITY: Rockville
STATE: Maryland
COUNTRY: USA
ZIP: 20850
COMPUTER READABLE FORM:
MEDIUM TYPE: Diskette, 3.50 inch, 1.4mb storage
COMPUTER: HP Vectra 486/33
OPERATING SYSTEM: MSDOS version 6.2
SOFTWARE: ASCII Text
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/961,083
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER:
FILING DATE:
ATTORNEY/AGENT INFORMATION:
NAME: Brookes, A. Anders
REGISTRATION NUMBER: 36,373
REFERENCE/DOCKET NUMBER: PB340P2
TELECOMMUNICATION INFORMATION:
TELEPHONE: (301) 309-8512
TELEFAX: (301) 309-8504
INFORMATION FOR SEQ ID NO: 162:
SEQUENCE CHARACTERISTICS:
LENGTH: 223 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: protein
US-08-961-083-162
Query Match 12.7%; Score 137; DB 3; Length 223;
Best local Similarity 30.8%; Pred. No. 8.8e-07;
Matches 45; Conservative 23; Mismatches 62; Indels 16; Gaps 7;
QY 64 QIRKDKSKVAGYIEFDADIKERYPGPATPEQNLNGVSFAENE-SLDQDNIS--IAGH 120
DB 52 QVSDPDVAYGYLISPLSEIMEPVYLG-ADYHHLGMLAHVDETPLPLDGTGIRSYIAGH 110


```

: TITLE OF INVENTION: STREPTOCOCCUS PNEUMONIAE 37-KDA SURFACE
: TITLE OF INVENTION: ADHESION A PROTEIN
: NUMBER OF SEQUENCES: 4
: CORRESPONDENCE ADDRESS:
: ADDRESS: Needle & Rosenberg, P.C.
: STREET: 127 Peachtree Street, Suite 1200
: CITY: Atlanta
: STATE: Georgia
: COUNTRY: USA
: ZIP: 30303-1811
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: Patentin Release #1.0, Version #1.30
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/715,131
: FILING DATE:
: CLASSIFICATION: 435
: ATTORNEY/AGENT INFORMATION:
: NAME: Spratt, Gwendolyn D.
: REGISTRATION NUMBER: 36,016
: REFERENCE/DOCKET NUMBER: 14114.0200
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: (404) 688-0770
: TELEFAX: (404) 688-9880
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 309 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: US-08-715-131-2

```

```

Query Match      8.2%; Score 88.5; DB 2; Length 309;
Best Local Similarity 28.0%; Pred. No. 0.18;
Matches 49; Conservative 18; Mismatches 53; Indels 55; Gaps 11;

QY 23 LFAKPHIDNYLHDKDEKIEYDKNVKEQASK-----DKKQAK-POIPKDK-----69
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 149 IRK-NIAKQLSAKDKPNK-EYEKMLKEYTDKLDKDESKDKPKITPAEKKLIYTSQG 206

QY 70 -----SKVAGYIEIPDADIKPEYPPAPPEQNLRGVS-----FAEENESLDDQ-113
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 207 AFKFEKAVG--VPSAYIWEINTEEGTEPEQIKTLVEKLRQTKVPSLFE--SSVDDRP 261

QY 114 -----NISTAGHTFIDRPYQFTNLKAKKGSWYFEKVGNETRYKMTSIRD 160
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 262 MKTVSODTNPIYAQIFTD-----SIAEGKEGDSY-----SMKKYNLDKIAE 305

```

```

RESULT 5
: US-08-922-837-2
: Sequence 2, Application US/08922837
: Patent No. 5888770
: GENERAL INFORMATION:
: APPLICANT: Chalker, Allison F.
: APPLICANT: Felin, Maria M.Z.
: APPLICANT: Brown, James R.
: APPLICANT: Bryant, Alexander P.
: TITLE OF INVENTION: No. 5888770el SpoilIE
: NUMBER OF SEQUENCES: 2
: CORRESPONDENCE ADDRESS:
: ADDRESSEE: Dechert, Price & Rhoads
: STREET: 4000 Bell Atlantic Tower, 1717 Arch Stre
: CITY: Philadelphia
: STATE: PA
: COUNTRY: USA
: ZIP: 19103-2793
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Diskette
: COMPUTER: IBM Compatible

```

```

: OPERATING SYSTEM: DOS
: SOFTWARE: FastSeq for Windows Version 2.0
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/922,837
: FILING DATE:
: CLASSIFICATION: 435
: PRIOR APPLICATION NUMBER:
: APPLICATION INFORMATION:
: FILING DATE:
: ATTORNEY/AGENT INFORMATION:
: NAME: Dickinson, O. Todd
: REGISTRATION NUMBER: 28,354
: REFERENCE/DOCKET NUMBER: GM10077
: TELECOMMUNICATION INFORMATION:
: TELEPHONE: 215/994-2252
: TELEFAX: 215/994-2222
: TELEX:
: INFORMATION FOR SEQ ID NO: 2:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 783 amino acids
: TYPE: amino acid
: STRANDEDNESS: single
: TOPOLOGY: linear
: US-08-922-837-2

```

```

Query Match      7.7%; Score 82.5; DB 2; Length 783;
Best Local Similarity 17.2%; Pred. No. 2.9;
Matches 42; Conservative 43; Mismatches 56; Indels 103; Gaps 10;

QY 11 IAGVILVAAYLFAKPHIDNYL-----HD 35
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 156 LIGVALYIPAFLEFS--NIGTYFTGISLILVGSILVSPMSGVYDIAEFPSGFAKMEGHE 213

QY 36 KDKDEKIEYDKNVKEQASKDKQKQAKPOIPKRSKAVAGIIEIPDADI-----KPYV 88
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 214 RRKEERFVQGEKAKQAKAEKAR-----LEQETTERA-LLDLPVDMETGETILTEEAVQ 266

QY 89 PGATPEQL-----NGVSFAEENESLDDONISIAHTFIDRPNYOFTNLKAKKG 139
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 267 NLPIPEEKVPEPILIPQELKFPEDSDSDDEDVY-----DFS-----307

QY 140 SMVYFVGNETRYKMTSIRDVPRPDVGLDEQKQKDKQTLTTCDDYNKGTGWKERRKT 199
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 308 -----AKALEYKLPQLFAP-----DKPRDQSK-----EKKIVRENIRKI 343

QY 200 FVAT 203
    |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||  |||
DB 344 LEAT 347

```

```

RESULT 6
: US-09-196-857-2
: Sequence 2, Application US/09196857A
: Patent No. 6100069
: GENERAL INFORMATION:
: APPLICANT: Zalacain, Magdalena
: APPLICANT: Brown, James R.
: APPLICANT: Biswas, Sanjoy
: APPLICANT: Thirup, John P.
: APPLICANT: Lawlor, Elizabeth J.
: APPLICANT: Mooney, Jeffrey
: APPLICANT: Zhong, Yiyi
: APPLICANT: Dehouck, Christine
: APPLICANT: Jaworski, Deborah D.
: APPLICANT: Wang, Min
: APPLICANT: Warren, Richard L.
: APPLICANT: Schilling, Lisa K.
: TITLE OF INVENTION: No. 6100069el t1g
: FILE REFERENCE: GM10118
: CURRENT APPLICATION NUMBER: US/09/196,857A
: CURRENT FILING DATE: 1998-11-20
: EARLIER APPLICATION NUMBER: 60/082,418

```

EARLIER FILING DATE: 1997-11-20
NUMBER OF SEQ ID NOS: 5
SOFTWARE: FastSeq for Windows Version 3.0
SEQ ID NO 2
LENGTH: 427
TYPE: PRF
ORGANISM: Streptococcus pneumoniae
US-09-196-857-2

Query Match 7.64; Score 82; DB 3; Length 427;
Best Local Similarity 21.88; Pred. NO. 1.4;
Matches 43; Conservative 30; Mismatches 70; Indels 54; Gaps 8;

OY 18 LVAAYLFAPKPH-----IDNYLHDKDKDEKIEQYDKNVKPOASKDKKQAKP 63
DB 105 VIAAEVYTKPEVKLGDKYKNLEVSVDVEKEVTDADVEERIEERNNTLTLYLKEAAEAGND 164
OY 64 QIPDKSKVAGYIEIPADIKPEVYPGPATPEQLNRGVSAEENESLDDQNISAGH-- 120
DB 165 TVVID--FVGSIGVERD-----GKGGENFSLGSGQFIPIGFEDQ--LVGHSAG 210
OY 121 -----TFIDRPVYQFTNKAARKGSMVYFKVGNETKRYKMTSIRDYKPTPDVGYLDEOK 173
DB 211 ETVDVIVTEPE--DYQAEGL-AGKEAFV-----TTIHEVKAKEVPAIDDEL 254
OY 174 GKDKQLITICDDYNEK 190
DB 255 AKDIDEVEETLADLKEK 271

RESULT 7

US-08-285-440-6
Sequence 6, Application US/08285440
Patent No. 5532337

GENERAL INFORMATION:

APPLICANT: Ken Ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 5.25 inch, 500 kb
COMPUTER: IBM Compatible
OPERATING SYSTEM: MS-DOS
SOFTWARE: Displaywrite

CURRENT APPLICATION DATA:

APPLICATION NUMBER: US/08/285.440

FILING DATE:

CLASSIFICATION: 536

PRIOR APPLICATION DATA:

APPLICATION NUMBER: 07/858,947

FILING DATE: March 27, 1992

ATTORNEY/AGENT INFORMATION:

NAME: Warren M. Cheek, Jr.

REGISTRATION NUMBER: 33,367

REFERENCE/DOCKET NUMBER:

TELECOMMUNICATION INFORMATION:

TELEPHONE: 202-371-8850

TELEFAX:

TELEX:

INFORMATION FOR SEQ ID NO: 6:

SEQUENCE CHARACTERISTICS:

LENGTH: 558 amino acids

TYPE: amino acid

STRANDEDNESS: single

TOPOLOGY: linear

MOLECULE TYPE: peptide

HYPOTHETICAL:
ANTI-SENSE:
FRAGMENT TYPE:
ORIGINAL SOURCE:
ORGANISM:

STRAIN:

INDIVIDUAL ISOLATE:
DEVELOPMENTAL STAGE:

HAPLOTYPE:

TISSUE TYPE:

CELL LINE:

ORGANELLE:

IMMEDIATE SOURCE:

LIBRARY:

CLONE:

POSITION IN GENOME:

CHROMOSOME/SEGMENT:

MAP POSITION:

UNITS:

FEATURE:

NAME/KEY:

LOCATION:

IDENTIFICATION METHOD:

OTHER INFORMATION:

PUBLICATION INFORMATION:

AUTHORS:

TITLE:

JOURNAL:

VOLUME:

ISSUE:

PAGES:

DATE:

DOCUMENT NUMBER:

FILING DATE:

PUBLICATION DATE:

RELEVANT RESIDUES IN SEQ ID NO:

US-08-285-440-6

Query Match 7.64; Score 82; DB 1; Length 558;

Best Local Similarity 25.94; Pred. NO. 2;

Matches 29; Conservative 16; Mismatches 43; Indels 24; Gaps 4;

OY 30 DNYLHDKDKDEKIEQ-----YDKNVKQASDKKQAKPQIPKSKVAGYTE 77

DB 176 ENKKEDKEKEEBEKKPKRGSGIGENQEEKTKYQAKREKLEQEDKPTKKE-----E 227

OY 78 IPDADIKPEVYPGPATPEQLNRGVSAE--ENESLDDQNISAGHFTIDRP 126

DB 228 IKDEIKKKDKPEKEVKSFMDRKKGFTLVKSGNGEFTMHKLKHTNFT-SRP 278

RESULT 8

US-08-630-349-6

Sequence 6, Application US/08630349
Patent No. 5739008

GENERAL INFORMATION:

APPLICANT: Ken Ichiro HAYASHI et al.
TITLE OF INVENTION: POLYPEPTIDE
NUMBER OF SEQUENCES: 32

CORRESPONDENCE ADDRESS:

ADDRESSEE: Wenderoth, Lind & Ponack
STREET: 805 Fifteenth Street, N.W., #700
CITY: Washington
STATE: D.C.
COUNTRY: U.S.A.
ZIP: 20005

COMPUTER READABLE FORM:

MEDIUM TYPE: Diskette, 3.5 inch, 144 mb

COMPUTER: IBM Compatible

OPERATING SYSTEM: MS-DOS

SOFTWARE: Wordperfect 5.1

```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/630,349
FILING DATE: April 10, 1996
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 08/285,440
FILING DATE: August 4, 1994
PRIOR APPLICATION DATA:
APPLICATION NUMBER: 07/858,947
FILING DATE: March 27, 1992
ATTORNEY/AGENT INFORMATION:
NAME: Warren M. Cheek, Jr.
REGISTRATION NUMBER: 33,367
INFORMATION FOR SEQ ID NO: 6:
SEQUENCE CHARACTERISTICS:
LENGTH: 558 amino acids
TYPE: amino acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: peptide
US-08-630-349-6

```

```

Query Match          7.6%; Score 82; DB 1; Length 558;
Best Local Similarity 25.9%; Pred. No. 2;
Matches 29; Conservative 16; Mismatches 43; Indels 24; Gaps 4;

QY 30 DNVLDHDKDEKIEQ-----YDKNVKQASKDKKQAKPOIPKDKSVAGIYE 77
DB 176 ENKKKEKEKEEKEEKKPKSGENGEKGTVKQAKREKLODKPTFKKE-----E 227
QY 78 IEDADIKPEVYPGATPEQLNKGVSFAE--ENESLDDONISAGHTFIDRP 126
DB 228 IKDEKIKKKEPEEYKSFMDRKKGFTEVKSQNGEMFTIKLKTENTF-SRP 278

```

```

RESULT 9
US-08-676-967-1
Sequence 1, Application US/08676967
Patent No. 5747317
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,967
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant

```

```

MOLECULE TYPE: peptide
US-08-676-967-1
Query Match          7.5%; Score 80.5; DB 1; Length 759;
Best Local Similarity 21.6%; Pred. No. 4.5;
Matches 35; Conservative 32; Mismatches 80; Indels 15; Gaps 4;

QY 35 DKDKDEKIEQYDKNVKEQASKDKKQAKPOIPKDKSVAGIETIPADIKPEVYPGATP 94
DB 227 DMEEEENDDDDDDDEEDGDFDEDEEENI---ESVYTPVQIKRAVNR---PAPAKS 280
QY 95 EQLNRGVSAEENESLDDONISAGHTFIDRPVQFTNLKAAGKSNVFRVGNETPKYK 154
DB 281 SDHSEEDSDLEEDSDIDGELAQSPTSTEEQEDKAVQSNKKKR-----KLPSDVNEGK 335
QY 155 MTSIR---DVKPTDGVLDQKQKQKQLTLITCDYNETG 192
DB 336 TVEIRNLSPDEEELGELLQDFGELKYRIVLHPDTEHSGK 377

```

```

RESULT 10
US-08-676-974-1
Sequence 1, Application US/08676974
Patent No. 5770422
GENERAL INFORMATION:
APPLICANT: COLLINS, KATHLEEN
TITLE OF INVENTION: Human Telomerase
NUMBER OF SEQUENCES: 10
CORRESPONDENCE ADDRESS:
ADDRESS: Science & Technology Law Group
STREET: 268 Bush Street, Suite 3200
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94104
COMPUTER READABLE FORM:
MEDIUM TYPE: floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patent Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/676,974
FILING DATE:
CLASSIFICATION: 530
ATTORNEY/AGENT INFORMATION:
NAME: Osman Ph.D., Richard A
REGISTRATION NUMBER: 36,627
REFERENCE/DOCKET NUMBER: UCB96-055
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415)343-4341
TELEFAX: (415)343-4342
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 759 amino acids
TYPE: amino acid
STRANDEDNESS:
TOPOLOGY: not relevant
MOLECULE TYPE: peptide
US-08-676-974-1
Query Match          7.5%; Score 80.5; DB 1; Length 759;
Best Local Similarity 21.6%; Pred. No. 4.5;
Matches 35; Conservative 32; Mismatches 80; Indels 15; Gaps 4;

QY 35 DKDKDEKIEQYDKNVKEQASKDKKQAKPOIPKDKSVAGIETIPADIKPEVYPGATP 94
DB 227 DMEEEENDDDDDDDEEDGDFDEDEEENI---ESVYTPVQIKRAVNR---PAPAKS 280
QY 95 EQLNRGVSAEENESLDDONISAGHTFIDRPVQFTNLKAAGKSNVFRVGNETPKYK 154
DB 281 SDHSEEDSDLEEDSDIDGELAQSPTSTEEQEDKAVQSNKKKR-----KLPSDVNEGK 335

```



```

; INFORMATION FOR SEQ ID NO: 5:
; SEQUENCE CHARACTERISTICS:
; LENGTH: 273 amino acids
; TYPE: amino acid
; TOPOLOGY: linear
; MOLECULE TYPE: protein
US-08-137-175A-5

```

```

Query Match      7.48; Score 80; DB 1; Length 273;
Best Local Similarity 22.68; Pred. No. 1.2;
Matches 54; Conservative 39; Mismatches 84; Indels 62; Gaps 12;

```

```

QY 1 MKWTNRLMTIAGVYLIVAAVLFAPKPHIDN-----YLDKDKD--- 39
   |||      : |||  | : | : :      : |||
Db 1 MKKY-----LIGIGLIL--ALIAKQWVSSIDKNSASVDLPGEKVLVSKEDKDGKY 52
   |||      : |||  | : | : :      : |||
QY 40 -----EKIEQYDKNVKEQAS-----KDKKQOAKPQIPKDKSVAGYIEIPDADIKEPV 87
   |||      : |||  | : | : :      : |||
Db 53 SLKATVDKIELKGTSDKNGSGVLEGTQDKSKAKLTADLSKTT--FELFKEDGKTLV 110
   |||      : |||  | : | : :      : |||
QY 88 YPGPATPEQLNRGVSAFENESLDDQNISIAHTEFIDRPNYQFTNLKAARKGS---MAY 143
   : : : : : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
Db 111 SRVSSKOKTSTDEMFNEKE-----LSAKTMTRENGTKLEYTEMKSDGTGKAKEVLKN 164
   : : : : : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
QY 144 FKVGNETRKRKYMTSIRDYKPTDVGVLDEQKGRKDKOLITICDDYN----EKTGVWEKR 197
   | : : : : | : | : : : : : : : : : : : : : : : : : : : : : : : : :
Db 165 FTLEGKAVANDKVT--LEVKEGTV--TLSEKELAKSGEVT--VALNDPTNTQATKKTGAMDSK 219
   | : : : : | : | : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Search completed: March 2, 2001, 10:23:05
Job time: 87 sec

```


GenCore version 4.5
Copyright (c) 1993 - 2000 CompuGen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:20:33 ; Search time 16.27 Seconds
(without alignments)
432.939 Million cell updates/sec

Title: US-09-292-437-3
Perfect score: 1076
Sequence: 1 MKRWNLMTIAGVLLVLA.....YNEKGVEMERKIFVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 268485 seqs, 34193795 residues
Total number of hits satisfying chosen parameters: 268485

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

- Database :
- 1: /SIDS1/gcgdata/geneseq/geneseqp/AA1980.DAT:**
 - 2: /SIDS1/gcgdata/geneseq/geneseqp/AA1981.DAT:**
 - 3: /SIDS1/gcgdata/geneseq/geneseqp/AA1982.DAT:**
 - 4: /SIDS1/gcgdata/geneseq/geneseqp/AA1983.DAT:**
 - 5: /SIDS1/gcgdata/geneseq/geneseqp/AA1984.DAT:**
 - 6: /SIDS1/gcgdata/geneseq/geneseqp/AA1985.DAT:**
 - 7: /SIDS1/gcgdata/geneseq/geneseqp/AA1986.DAT:**
 - 8: /SIDS1/gcgdata/geneseq/geneseqp/AA1987.DAT:**
 - 9: /SIDS1/gcgdata/geneseq/geneseqp/AA1988.DAT:**
 - 10: /SIDS1/gcgdata/geneseq/geneseqp/AA1989.DAT:**
 - 11: /SIDS1/gcgdata/geneseq/geneseqp/AA1990.DAT:**
 - 12: /SIDS1/gcgdata/geneseq/geneseqp/AA1991.DAT:**
 - 13: /SIDS1/gcgdata/geneseq/geneseqp/AA1992.DAT:**
 - 14: /SIDS1/gcgdata/geneseq/geneseqp/AA1993.DAT:**
 - 15: /SIDS1/gcgdata/geneseq/geneseqp/AA1994.DAT:**
 - 16: /SIDS1/gcgdata/geneseq/geneseqp/AA1995.DAT:**
 - 17: /SIDS1/gcgdata/geneseq/geneseqp/AA1996.DAT:**
 - 18: /SIDS1/gcgdata/geneseq/geneseqp/AA1997.DAT:**
 - 19: /SIDS1/gcgdata/geneseq/geneseqp/AA1998.DAT:**
 - 20: /SIDS1/gcgdata/geneseq/geneseqp/AA1999.DAT:**
 - 21: /SIDS1/gcgdata/geneseq/geneseqp/AA2000.DAT:**

Pred. NO. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	160.5	14.9	251	Y00252	Enterococcus faeca
2	158.5	14.7	284	Y00158	Enterococcus faeca
3	154.5	14.4	197	Y00253	Enterococcus faeca
4	146	13.6	236	Y00159	Enterococcus faeca
5	137	12.7	223	W61218	Streptococcus pneu
6	90	8.4	167	W89768	Streptococcus pneu
7	90	8.4	167	Y20763	Human neurofilamen
8	89.5	8.3	411	W68011	Yeast immunophilin
9	89.5	8.3	1617	Y35617	Chlamydia pneumonia
10	88.5	8.3	359	W98324	H. pylori GPO 141
11	88.5	8.2	289	W55072	Streptococcus pneu
12	88.5	8.2	309	Y30350	37 kDa pneumococca

13	88.5	8.2	309	20	W82496
14	88.5	8.2	310	21	Y81668
15	87	8.1	310	20	Y00215
16	87	8.1	347	10	Y00214
17	87	8.1	493	13	R26944
18	84.5	7.9	293	20	Y00257
19	84.5	7.9	316	20	Y00256
20	84.5	7.8	170	21	Y84722
21	83.5	7.8	680	12	R12540
22	83.5	7.8	2432	21	R55565
23	83	7.7	1035	20	W90264
24	82.5	7.7	768	21	Y81596
25	82.5	7.7	783	20	W97416
26	82.5	7.7	991	21	Y83171
27	82.5	7.7	991	21	Y70120
28	82	7.6	235	19	W62784
29	82	7.6	235	19	W62787
30	82	7.6	235	19	W62788
31	82	7.6	280	10	P90366
32	82	7.6	307	19	W80703
33	82	7.6	427	20	Y16590
34	82	7.6	558	13	R27363
35	82	7.6	911	20	Y55957
36	82	7.6	1897	21	Y81785
37	82	7.6	1897	21	Y56100
38	81.5	7.5	783	11	R05804
39	81	7.5	289	20	Y38690
40	80.5	7.5	653	20	Y35021
41	80.5	7.5	759	19	W41927
42	80.5	7.5	759	19	W46593
43	80	7.4	173	18	W28351
44	80	7.4	186	19	W77774
45	80	7.4	273	12	R13141

ALIGNMENTS

RESULT	ID	Y00252 standard; Protein: 251 AA.	ALIGNMENTS
Y00252	Y00252	standard; Protein: 251 AA.	
AC	Y00252		
XX	20-APR-1999	(first entry)	
DE	Enterococcus faecalis	protein EFL30.	
XX	Enterococcus faecalis	infection; vaccine; immune response; diagnosis;	
KW	detection; attenuation; antigenic.		
OS	Enterococcus faecalis.		
XX	W09850554-A2.		
PN	12-NOV-1998.		
PD	04-MAY-1998;	98WO-US08959.	
XX	14-NOV-1997;	97US-0066009.	
PR	06-MAY-1997;	97US-0044031.	
PR	16-MAY-1997;	97US-0046655.	
XX	(HUMA-) HUMAN GENOME SCI INC.		
PA	Bailey C, Choi GH, Hromockyj A, Kunsch CA;		
PI	WPI: 1999-070095/06.		
XX	N-PSDB; X20242.		
DR	New isolated Enterococcus faecalis polynucleotides - used to develop		
XX	products for the detection of Enterococcus and for use in vaccines		
PT	for prevention or attenuation of Enterococcus infection		
PT			

XX Claim 9; Page 241; 301pp; English.

PS The present sequence represents an antigenic polypeptide fragment

CC isolated from *Enterococcus faecalis*. The present invention describes

CC genes, proteins and antigenic polypeptides isolated from *E. faecalis*.

CC The proteins can be used in vaccines for preventing or attenuating an

CC infection caused by a member of the *Enterococcus* genus in an animal.

CC They can also be used for detecting *Enterococcus* antibodies in a sample.

CC The nucleotide sequences can be used for detecting *Enterococcus* nucleic

CC acids. Products from the present invention can also be used for

CC screening compounds to identify agonists and antagonists of *E. faecalis*

CC protein activity.

CC Sequence 197 AA;

XX

XX

Query Match 14.4%; Score 154.5; DB 20; Length 197;

Best Local Similarity 27.6%; Pred. No. 1.9e-07;

Matches 40; Conservative 30; Mismatches 72; Indels 3; Gaps 3;

OY 62 KPQIKDKSKVAGYIEIPDADIKPEYVGPATPEQNLNGVSFAENE--SLDDONISIAHGTFTDRN 121

DB 34 kqfemknlpyigaiaipsveinlpitkylsnvailt-gagtmkedqvmgkmyalashr 92

OY 122 FIDRENYOFTNLKAARKGSMYFVKVNETRKYKMTSIRDVKEPTDVGVLDEOKKDKQLTL 181

DB 93 tedgysl-fsperlckkdeliyitdstvylyktsvkekleptrelldvpggn-mtcl 150

OY 182 ITCDYNEKTGVWEKRRKIFVATEVK 206

DB 151 itcgdlqatrtiavqgtlaatpik 175

RESULT 4

ID Y00159 standard; Protein; 256 AA.

XX

AC Y00159;

XX

DT 20-APR-1999 (first entry)

XX

DE Enterococcus faecalis antigenic polypeptide fragment EF079.

XX

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

KW detection; attenuation; antigenic.

XX

OS Enterococcus faecalis.

OS

XX

PN W09850554-A2.

XX

PD 12-NOV-1998.

XX

XX

PE 04-MAY-1998; 98WO-US08959.

XX

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;

XX

DR MPI: 1999-070095/06.

DR N-PSDB: X20149.

XX

XX

PT New isolated *Enterococcus faecalis* polynucleotides - used to develop

PT products for the detection of *Enterococcus* and for use in vaccines

PT for prevention or attenuation of *Enterococcus* infection

XX

PS Claim 9; Page 170; 301pp; English.

XX

XX The present sequence represents an antigenic polypeptide fragment

CC isolated from *Enterococcus faecalis*. The present invention describes

CC genes, proteins and antigenic polypeptides isolated from *E. faecalis*.

CC The proteins can be used in vaccines for preventing or attenuating an

CC infection caused by a member of the *Enterococcus* genus in an animal.

CC They can also be used for detecting *Enterococcus* antibodies in a sample.

CC The nucleotide sequences can be used for detecting *Enterococcus* nucleic

CC acids. Products from the present invention can also be used for

CC screening compounds to identify agonists and antagonists of *E. faecalis*

CC protein activity.

XX

XX

Sequence 256 AA;

XX

XX

Query Match 13.6%; Score 146; DB 20; Length 256;

Best Local Similarity 25.6%; Pred. No. 1.8e-06;

Matches 46; Conservative 37; Mismatches 83; Indels 14; Gaps 5;

OY 19 VAAVLEAKPHIDNYLHDKDKDEKIEQYDKNVKFA-----SKDKQQAQKQIPKDKS 70

DB 15 ilahyqakasgentkemaieqekmekngelakgsnpgldfsetqtktkpkdkasyfes 74

OY 71 KVAGYIEIPDADIKPEYVGPATPEQNLNGVSFAENE--SLDDONISIAHGTFTDRN 127

DB 75 htigvltipkinvtrlpif-dktnalllekysllegtsyptggtntchavisnrgjlpqak 133

OY 128 YQFTNLKAARKGSMYFVKVNETRKYKMTSIRDVKEPTDVGVLDEOKKDKQLTLITCDY 187

DB 134 l-ftldlpelkkdgyfievngkrlayqvqdklvepdtckdlhesgd-lyvllictcy 191

RESULT 5

ID W61218 standard; Protein; 223 AA.

XX

AC W61218;

XX

DT 02-OCT-1998 (first entry)

XX

DE Streptococcus pneumoniae SP0093 protein.

XX

KW Streptococcus pneumoniae; antigen; vaccine; infection; diagnosis;

KW detection; pneumonia; otitis media; meningitis.

XX

OS Streptococcus pneumoniae.

OS

XX

FT Key Location/Qualifiers

FT Misc-difference 33

FT /label= unknown

FT /note= "encoded by GNF"

XX

PN W09818930-A2.

XX

PD 07-MAY-1998.

XX

XX

PE 30-OCT-1997; 97WO-US19422.

XX

PR 31-OCT-1996; 96US-0029960.

XX

PA (HUMA-) HUMAN GENOME SCI INC.

XX

PI Choi GH, Hromockyj A, Johnson LS, Kunsch CA;

XX

DR MPI: 1998-272224/24.

DR N-PSDB: V27404.

XX

XX

PT Nucleic acid encoding antigenic peptide(s) from *Streptococcus*

PT pneumoniae - or their epitope-containing fragments, useful in

PT protective or therapeutic vaccines, and for diagnosis

XX

PS Claim 11; Page 82; 118pp; English.

XX

XX The present sequence represents a protein from *Streptococcus pneumoniae*.

CC The nucleic acid sequence encoding the *Streptococcus pneumoniae* protein

CC Can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.

XX Sequence 223 AA:

Query Match 12.7% Score 137; DB 19; Length 223;
 Best Local Similarity 30.8%; Pred. No. 1,1e-05;
 Matches 45; Conservative 23; Mismatches 62; Indels 16; Gaps 7;

QY 64 QIFRDKSKVAGYIEIPDADIKPEVPGAPPEOLNRGVSAEENE-SLDDONIS--IAGH 120
 Db 52 qvaddpavgyislpselempyig-adyhlgmgjahvdgfpplpldgqgirsyag 110
 QY 121 TFIIDRPVYOTNLKAAKSGMYFKVGNETRKRYKMTSIRBYKPPDVGVLDEQKGRKQLT 180
 Db 111 r-aepsnvtfrhldqjkhvgdalydngqelveyqymdteillpseweklesvsakn-lmt 168
 QY 181 LITCD--DYN-----EKTGWYER 196
 Db 169 litcdphtkrllyvferavayqk 194

RESULT 6
 W89768 standard; Protein; 167 AA.

XX AC W89768;
 XX DT 16-MAR-1999 (first entry)
 XX DE Staphylococcus aureus protein SEQ ID #5216.
 KW Computer readable medium; vaccine; S.aureus infection; immunodetection;
 KW cellulitis; eyelid infection; food poisoning; osteomyelitis; therapy;
 KW skin infection; surgical wound infection; scalded skin syndrome;
 KW toxic shock syndrome.
 XX OS Staphylococcus aureus.

XX EH Key Location/Qualifiers
 XX FT Misc-difference 1..16 /note="these residues represent a line of missing text
 XX FT in the sequence listing in the specification.
 XX FT They are included to maintain the residue
 XX FT numbering given in the specification for this
 XX FT protein sequence"

XX PN EP786519-A2.
 XX PD 30-JUL-1997.
 XX PF 07-JAN-1997; 97EP-0100117.
 XX PR 05-JAN-1996; 96US-0009861.
 XX PA (HUMA-) HUMAN GENOME SCI INC.
 XX PI Barash SC, Choi GH, Dillon PJ, Fannon MR, Kunsch CA;
 XX PI Rosen CA;
 XX DR WPI; 1997-374922/35.
 XX XX

PT Polynucleotide(s) and proteins derived from Staphylococcus aureus
 PT stored on computer readable medium and used in the production of
 PT anti-S.aureus vaccines
 XX
 PS Claim 23: Page 3213-3214; 3271pp; English.

XX This sequence represents a Staphylococcus aureus protein sequence of the
 CC invention. The DNA sequences encoding the S.aureus proteins are recorded
 CC on a computer readable medium, preferably selected from a floppy or hard
 CC disk, random access memory (RAM), read-only memory (ROM) or CD-ROM.
 CC Homology searches using the S.aureus DNA sequences allows putative
 CC functions to be assigned so that protein-encoding or regulatory regions
 CC of commercial, therapeutic or industrial importance can be obtained.
 CC Specifically, sequences which are likely to encode antigens have been
 CC identified and these polypeptides can be used in a vaccine composition
 CC against S.aureus infection. The polypeptides can also be used in a kit
 CC for the immunodetection of S.aureus in a sample. S.aureus is implicated
 CC in numerous human diseases, including cellulitis, eyelid infections, food
 CC poisoning, osteomyelitis, skin and surgical wound infections, scalded
 CC skin syndrome, toxic shock syndrome, etc. Organisms transformed with the
 CC DNA sequences can be used for recombinant production of the polypeptides.
 CC The new DNA sequences (and their fragments) are useful as primers or
 CC probes for isolating homologues of any of the 5191 S.aureus DNA sequences
 CC contained on the computer readable medium.

XX Sequence 167 AA;

Query Match 8.4% Score 90; DB 18; Length 167;
 Best Local Similarity 20.8%; Pred. No. 0.26;
 Matches 40; Conservative 30; Mismatches 58; Indels 64; Gaps 4;

QY 5 TNRLMTIAGVLIIVAAVYLFKAPHIDNYLIHDKD---DEKIEQDKNKVEQASRD--KKQ 59
 Db 28 takvyslatvllllyglvafalfayvdhsknkemlnqkqgqkxkqknaekerkkq 87
 QY 60 QAKPQIPKSKVAGYIEIPDADIKPEVPGAPPEOLNRGVSAEENESLDDONISIAG 119
 Db 88 qeekegneldsqangyqqlpqngnygyvppqgqptk----- 124
 QY 120 HTFIIDRPVYOTNLKAAKSGMYFKVGNETRKRYKMTSIRBYKPPDVGVLDEQKGRKQLT 179
 Db 125 ----qrpakeenddkask-----deskdodka 148
 QY 180 TLITCDYNEKT 191
 Db 149 sqkdsdnqkkt 160

RESULT 7
 Y20763
 ID Y20763 standard; Protein; 554 AA.

XX AC Y20763;
 XX DT 22-JUL-1999 (first entry)
 XX DE Human neurofilament-M mutant protein fragment 45.

XX KW Human; beta-amyloid precursor protein; beta-APP; diagnosis; cancer;
 KW frameshift mutation; age-related disease; neurodegenerative disorder;
 KW Alzheimer's disease; Down's syndrome; myotonic dystrophy; neuronal;
 KW Huntington's disease; multiple sclerosis; alcoholic liver disease;
 KW diabetes mellitus type II; microtubule associated protein; Tau; Big Tau;
 KW ubiquitin B; apolipoprotein E; MAP2; neurofilament-L; neurofilament-M;
 KW neurofilament-F; presenilin I; presenilin II; cellular tumour antigen;
 KW glial fibrillary acidic protein; GFAP; p53; semaphorin III; HNP-1;
 KW bcl-2; B-cell leukemia/lymphoma 2 proto-oncogene; HMGP-C; NSP-A;
 KW high mobility group protein-C; neuroendocrine specific protein A.

XX OS Synthetic.
 XX OS Homo sapiens.
 XX XX

DR WPI: 1998-272224/24.
 DR N-PSDB: V27333.
 XX Nucleic acid encoding antigenic peptide(s) from Streptococcus
 PT pneumoniae - or their epitope-containing fragments, useful in
 PT protective or therapeutic vaccines, and for diagnosis
 XX
 PS Claim 11; Page 52; 118pp; English.
 XX
 CC The present sequence represents a protein from Streptococcus pneumoniae.
 CC The nucleic acid sequence encoding the streptococcus pneumoniae protein
 CC can be useful in vaccines for inducing protective antibodies against
 CC Streptococcus pneumoniae, for treatment or prevention of infection e.g.
 CC pneumonia, otitis media or meningitis. Probes based on the nucleic acid
 CC are used to detect Streptococcus infection (by usual hybridisation or
 CC amplification methods), also for isolating Streptococcus genes or their
 CC allelic variants. The protein can be used similarly to detect specific
 CC antibodies in standard immunoassays, especially for diagnosing or
 CC monitoring infections. Antibodies which bind the protein are used to
 CC detect corresponding antigens, to purify the protein and for passive
 CC immunisation (optionally coupled to a toxin). Vaccines are administered,
 CC e.g. by injection, orally or through the skin, typically at 0.01-1000
 CC (especially 10-300) mu g/ml per dose.
 CC
 SQ Sequence 289 AA:
 Query Match 8.2%; Score 88.5; DB 19; Length 289;
 Best Local Similarity 28.0%; Pred. No. 0.8; Mismatches 53; Indels 55; Gaps 11;
 Matches 49; Conservative 18;
 QY 23 LFAKPHIDNVLHDKDKDEKIEQYDKNVKEQASK-----DKKOAK-POIPKDK----- 69
 Db 129 ifak-nlakqlsakdpnk-efyeknlkeytdklldkeskdkfnkipekklivtseg 186
 QY 70 -----SKVAGYIEIPDADIKPEYVPGPAPPEQNLNGVS-----FAENESLDDQ- 113
 Db 187 afkyfskayg---vpsaylweinteegtpedqiklveklrqlkypslive--ssvddrp 241
 QY 114 -----NISAGHTEFIDRPNYQFTNLKAAGKSMWYFVGNETRRKYMSTIRD 160
 Db 242 mktvsgdtnlpyaqiftd-----siaegkgdsdy-----smmkynldkiaae 285
 RESULT 12
 Y30350
 ID Y30350 standard; Protein: 309 AA.
 XX Y30350;
 AC
 XX
 DT 09-NOV-1999 (first entry)
 XX
 DE 37 kDa pneumococcal surface adhesion A protein (PsaA).
 XX
 KW Pneumococcal surface adhesion A protein; PsaA: monoclonal antibody;
 KW vaccine; Streptococcus pneumoniae infection.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN W09945121-A1.
 XX
 PD 10-SEP-1999.
 XX
 PF 26-FEB-1999; 99WO-US04326.
 XX
 PR 02-MAR-1998; 98US-0076565.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Ades EW, Carlone GM, Sampson JS, Tharpe JA, Westerink MAJ;
 PI Zeiler JL;
 PS
 DR WPI: 1999-540849/45.

DR N-PSDB: Z10411.
 XX
 PT New peptides corresponding to Streptococcus pneumoniae PsaA, used
 PT for treating or preventing Streptococcus pneumoniae infection in a
 PT subject
 XX
 PS Example 8; Page 53-54; 58pp; English.
 XX
 CC The present sequence represents a pneumococcal surface adhesion A protein
 CC (PsaA). The specification describes monoclonal antibodies which bind
 CC epitopes of the PsaA protein (see Y30351-54). These peptides can be
 CC used in vaccines to prevent Streptococcus pneumoniae infections. The
 CC antibodies of the invention can also be used to detect S. pneumoniae in
 CC a sample or individual.
 CC
 SQ Sequence 309 AA:
 Query Match 8.2%; Score 88.5; DB 20; Length 309;
 Best Local Similarity 28.0%; Pred. No. 0.88; Mismatches 53; Indels 55; Gaps 11;
 Matches 49; Conservative 18;
 QY 23 LFAKPHIDNVLHDKDKDEKIEQYDKNVKEQASK-----DKKOAK-POIPKDK----- 69
 Db 149 ifak-nlakqlsakdpnk-efyeknlkeytdklldkeskdkfnkipekklivtseg 206
 QY 70 -----SKVAGYIEIPDADIKPEYVPGPAPPEQNLNGVS-----FAENESLDDQ- 113
 Db 207 afkyfskayg---vpsaylweinteegtpedqiklveklrqlkypslive--ssvddrp 261
 QY 114 -----NISAGHTEFIDRPNYQFTNLKAAGKSMWYFVGNETRRKYMSTIRD 160
 Db 262 mktvsgdtnlpyaqiftd-----siaegkgdsdy-----smmkynldkiaae 305
 RESULT 13
 W82496
 ID W82496 standard; Protein: 309 AA.
 XX W82496;
 AC
 XX
 DT 04-MAR-1999 (first entry)
 XX
 DE S. pneumoniae 37-kDa surface adhesion A protein.
 XX
 KW Surface adhesion A protein; vaccine; detection; serotype; antibody;
 KW diagnostic; immunoassay; treatment; infection; anti-idiotypic.
 XX
 OS Streptococcus pneumoniae.
 XX
 PN US5854416-A.
 XX
 PD 29-DEC-1998.
 XX
 PF 17-SEP-1996; 96US-0715131.
 XX
 PR 17-SEP-1996; 96US-0715131.
 PR 17-SEP-1991; 91US-0791377.
 PR 04-APR-1994; 94US-0222179.
 XX
 XX (USSH) US DEPT HEALTH & HUMAN SERVICES.
 PA Ades EW, Carlone GM, Russell H, Sampson JS, Tharpe JA;
 PI N-PSDB: V73914.
 DR WPI: 1999-095007/08.
 DR N-PSDB: V73914.
 XX Nucleic acid encoding the 37 kDa surface adhesion A of Streptococcus
 PT pneumoniae - useful diagnostically and for production of
 PT recombinant polypeptides
 XX
 PS Claim 1; Column 33-34; 20pp; English.
 XX

CC This sequence represents a Streptococcus pneumoniae 37-kDa surface
 CC adhesion A protein. This encoding nucleic acid can be used in methods to
 CC express recombinant protein, as a source of primers for amplification (to
 CC identify and isolate related sequences, e.g. allelic variants) or probes
 CC for nucleic acid hybridisation tests for detecting S. pneumoniae, and in
 CC DNA vaccines. This protein and its fragments can be used to raise
 CC antibodies, in vaccines and for detecting S. pneumoniae (by reaction with
 CC specific antibodies). Antibodies are useful in diagnostic immunoassays,
 CC to treat infections and to raise anti-Idiotypic antibodies for use in
 CC vaccines. This protein is very highly conserved between the different
 CC serotypes of S. pneumoniae so is an excellent candidate for vaccine
 CC development.

XX Sequence 309 AA;

Query Match 8.2%; Score 88.5; DB 20; Length 309;

Best Local Similarity 28.0%; Pred. No. 0.88; Mismatches 53; Indels 55; Gaps 11;

Matches 49; Conservative 18; Mismatches 53; Indels 55; Gaps 11;

OY 23 LEAKPHIDNYLHDKDKDEKIEQYDKNVKEQASK---DKKQAK-POIPKDK-----69

DB 149 ifak-nlakqksakdpnk-efyeknlkeytdklidkldkeskdkfnkipaekklivtseg 206

OY 70 -----SKVAGYIEIPDADIKPEVYPGPATPPOLNRGVS-----FAEENESLDDQ- 113

DB 207 afkyfskayg---vpsayiwelnteegtlpeqiklveklrtqkypslfive--ssvddrp 261

OY 114 -----NISAGHTFIDRPNYOFTNLKAAKGSWYFKVGNERTKRYKMSIRD 160

DB 262 mktvsqdnipiyagiftd-----slaegqkegdsyy-----smmkynldkiaae 305

RESULT 14

ID Y81668 standard; Protein; 310 AA.

XX Y81668;

DT 24-MAY-2000 (first entry)

DE Streptococcus pneumoniae protein sequence ID205 - 4118.1.

KW Streptococcus pneumoniae; vaccine; screening; protein antigen;

KM antibacterial; antiinflammatory; meningitis; infection; diagnosis;

KW pneumococcal disease.

OS Streptococcus pneumoniae.

PN W0200006737-A2.

PD 10-FEB-2000.

PF 27-JUL-1999; 99WO-GB02451.

PR 27-JUL-1998; 98GB-0016337.

PR 19-MAR-1999; 99US-0125164.

PA (MICR-) MICROBIAL TECHNIQS LTD.

PI Gilbert CFG, Hansbro PM;

DR WPI: 2000-195300/17.

PT New Streptococcal protein, useful as a vaccine, for diagnosis of

PT pneumococcal diseases and for screening agents capable of antagonizing

PT or inhibiting expression of the protein

PS Claim 6; Page 90; 108pp; English.

XX Y81501 to Y81679 represent specifically claimed protein sequences

CC isolated from Streptococcus pneumoniae. A05407 to A05590 represent

CC specifically claimed nucleotide sequences isolated from S. pneumoniae.

CC The sequences have antibacterial and antiinflammatory properties.
 CC The protein sequences, and fragments of them, are useful as immunogens
 CC and/or antigens. The nucleotide sequences can be used in vaccines and in
 CC diagnostic assays. The proteins and nucleotides can be useful for the
 CC detection and diagnosis of S. pneumoniae. The protein sequences are also
 CC useful for screening an agent capable of antagonising, inhibiting or
 CC interfering with the function or expression of the proteins in which the
 CC agent is useful for treatment or prophylaxis of S. pneumoniae infection
 CC and meningitis. A05591 to A05614 represent primers used in the
 CC exemplification of the present invention.

XX Sequence 310 AA;

Query Match 8.2%; Score 88.5; DB 21; Length 310;

Best Local Similarity 28.0%; Pred. No. 0.88; Mismatches 53; Indels 55; Gaps 11;

Matches 49; Conservative 18; Mismatches 53; Indels 55; Gaps 11;

OY 23 LEAKPHIDNYLHDKDKDEKIEQYDKNVKEQASK---DKKQAK-POIPKDK-----69

DB 149 ifak-nlakqksakdpnk-efyeknlkeytdklidkldkeskdkfnkipaekklivtseg 206

OY 70 -----SKVAGYIEIPDADIKPEVYPGPATPPOLNRGVS-----FAEENESLDDQ- 113

DB 207 afkyfskayg---vpsayiwelnteegtlpeqiklveklrtqkypslfive--ssvddrp 261

OY 114 -----NISAGHTFIDRPNYOFTNLKAAKGSWYFKVGNERTKRYKMSIRD 160

DB 262 mktvsqdnipiyagiftd-----slaegqkegdsyy-----smmkynldkiaae 305

RESULT 15

ID Y00215 standard; Protein; 310 AA.

XX Y00215;

DT 20-APR-1999 (first entry)

DE Enterococcus faecalis antigenic polypeptide fragment EF106.

KW Enterococcus faecalis; infection; vaccine; immune response; diagnosis;

KM detection; attenuation; antigenic.

OS Enterococcus faecalis.

PN W09850554-A2.

PD 12-NOV-1998.

PF 04-MAY-1998; 98WO-US08959.

PR 14-NOV-1997; 97US-0066009.

PR 06-MAY-1997; 97US-0044031.

PR 16-MAY-1997; 97US-0046655.

PA (HUMA-) HUMAN GENOME SCI INC.

PI Bailey C, Choi GH, Hromockyj A, Kunsch CA;

DR WPI: 1999-070095/06.

DR N-PSDB; X20205.

PT New isolated Enterococcus faecalis polynucleotides - used to develop

PT products for the detection of Enterococcus and for use in vaccines

PT for prevention or attenuation of Enterococcus infection

PS Claim 9; Page 205; 301pp; English.

XX The present sequence represents an antigenic polypeptide fragment

CC isolated from Enterococcus faecalis. The present invention describes

CC genes, proteins and antigenic polypeptides isolated from E. faecalis.

CC The proteins can be used in vaccines for preventing or attenuating an

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 06:48:04 ; Search time 62.96 Seconds
(without alignments)
1589.589 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
Sequence: 1 atgaaaaaatggacaatcg.....tagctacagaagtaataaa 621

Scoring table: IDENTITY NUC
Gapop 10.0 , Gapext 1.0

Searched: 280836 seqs, 80580151 residues

Total number of hits satisfying chosen parameters: 561672

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database : Issued_Patents_NA.*

1: /cgn2_6/ptodata/1/lna/5A.COMB.seq:*
2: /cgn2_6/ptodata/1/lna/5B.COMB.seq:*
3: /cgn2_6/ptodata/1/lna/6.COMB.seq:*
4: /cgn2_6/ptodata/1/lna/PCrUS.COMB.seq:*
5: /cgn2_6/ptodata/1/lna/backfiles1.seq:*

Pred. NO. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
C 1	65.4	10.5	7218	1	US-08-232-463-14
2	59.8	9.6	3095	5	5231168-1
3	50.6	8.1	9636	1	US-08-323-170B-1
4	48.4	7.8	4248	3	US-08-678-614-1
5	48.2	7.8	1368	1	US-08-021-601-5
6	48.2	7.8	1368	1	US-08-082-849B-5
7	48.2	7.8	1368	4	PCr-US94-01624-5
8	48.2	7.8	1425	1	US-08-021-601-7
9	48.2	7.8	1425	1	US-08-082-849B-7
10	48.2	7.8	1425	4	PCr-US94-01624-7
11	48.2	7.8	1524	1	US-08-021-601-9
12	48.2	7.8	1524	1	US-08-082-849B-9
13	48.2	7.8	1524	4	PCr-US94-01624-9
14	48.2	7.8	3291	1	US-08-021-601-1
15	48.2	7.8	3291	1	US-08-082-849B-1
16	48.2	7.8	3291	4	PCr-US94-01624-1
17	46.2	7.4	2223	1	US-08-257-073-4
18	46.2	7.4	19124	2	US-08-487-826B-13
19	43	6.9	1360	3	US-08-961-083-37
C 20	42.6	6.9	1921	2	US-08-557-128-11
21	41.6	6.7	2042	2	US-08-933-821-16
22	41.6	6.7	2042	3	US-08-934-494-5
23	41.6	6.7	2042	3	US-08-960-507-16
24	41.6	6.7	2042	3	US-09-143-068-5
25	41.2	6.6	3892	3	US-08-569-214-1
26	41	6.6	1696	3	US-08-961-083-217
27	41	6.6	3022	3	US-08-961-083-215
28	40.8	6.6	4766	4	PCr-US93-07261-10

29	40.4	6.5	3498	3	US-08-293-728-1	Sequence 1, Appl
30	40.2	6.5	387	4	PCr-US96-05320A-1342	Sequence 1342, Ap
C 31	40.2	6.5	845	2	US-08-743-637B-25	Sequence 25, Appl
32	40.2	6.5	845	3	US-08-526-840B-25	Sequence 25, Appl
33	39.8	6.4	1098	3	US-09-248-335-35	Sequence 35, Appl
34	39.8	6.4	3665	2	US-08-682-517-13	Sequence 13, Appl
35	39.8	6.4	3665	2	US-08-682-517-14	Sequence 14, Appl
36	39.8	6.4	4197	2	US-08-682-517-14	Sequence 14, Appl
37	39.8	6.4	4197	2	US-08-682-517-7	Sequence 8, Appl
38	39.6	6.4	240	1	US-08-628-417-6	Sequence 8, Appl
39	39.6	6.4	1733	3	US-09-073-569-1	Sequence 1, Appl
40	39.6	6.4	4090	3	US-08-569-214-4	Sequence 4, Appl
41	38.4	6.2	2007	2	US-08-743-637B-169	Sequence 169, Appl
42	38.4	6.2	2007	3	US-08-526-840B-169	Sequence 169, App
43	38.2	6.2	6243	2	US-09-056-075-1	Sequence 1, Appl
C 44	38.2	6.2	8457	1	US-07-991-867B-1	Sequence 1, Appl
45	38.2	6.2	8457	2	US-08-544-332-1	Sequence 1, Appl

ALIGNMENTS

RESULT 1
US-08-232-463-14/C
Sequence 14, Application US/08232463
Patent No. 5670367
GENERAL INFORMATION:
APPLICANT: DORNER, F.
APPLICANT: SCHEIFLINGER, F.
APPLICANT: FALKNER, F. G.
TITLE OF INVENTION: RECOMBINANT FOWLPOX VIRUS
NUMBER OF SEQUENCES: 52
CORRESPONDENCE ADDRESS:
ADDRESSEE: Foley & Lardner
STREET: 1800 Diagonal Road, Suite 500
CITY: Alexandria
STATE: VA
COUNTRY: USA
ZIP: 22313-0299
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/232,463
FILING DATE:
CLASSIFICATION: 435
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US/07/935,313
FILING DATE:
APPLICATION NUMBER: EP 91 114 300.6
FILING DATE: 26-AUG-1991
ATTORNEY/AGENT INFORMATION:
NAME: BENT, Stephen A.
REGISTRATION NUMBER: 29,768
REFERENCE/DOCKET NUMBER: 30472/114 IMMU
TELECOMMUNICATION INFORMATION:
TELEPHONE: (703)836-9300
TELEFAX: (703)683-4109
TELEX: 899149
INFORMATION FOR SEQ ID NO: 14:
SEQUENCE CHARACTERISTICS:
LENGTH: 7218 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
IMMEDIATE SOURCE:
CLONE: PTZpT-FLS
US-08-232-463-14

Query Match 10.5%; Score 65.4; DB 1; Length 7218;

Best Local Similarity 5.38; Pred. No. 3.9e-07;
Matches 21; Conservative 225; Mismatches 151; Indels 0; Gaps 0;

[illegible]

```

RESULT 2
5231168-1
; Patent No. 5231168
; APPLICANT: DIEZGIEL, MORTEN;BORE, MARTIN;JEPSEN, SOREN
; VUDST, JENS;RIENECK, KLAUS;WIND, ANNETTE;JAKOBSEN, PALLE H.
; TITLE OF INVENTION: MALARIA ANTIGEN
; NUMBER OF SEQUENCES: 19
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/07/409,658
; FILING DATE: 18-SEP-1989
; SEQ ID NO:1:
; LENGTH: 3095
; 2231168-1

```

Query Match	9.6%;	Score 59.8;	DB 5;	Length 3095;
Best Local Similarity	44.4%;	Pred. No. 6.8e-06;		
Matches 241;	Conservative	0;	Mismatches 302;	Indels 0;
				Gaps 0;

QY 72 tgcTaaaccacatccgatataattcttcacggttaaagatcgtaaaagtgtgaac 131
|| || || || || || || || || || || || || || || || || || || || ||
Db 1260 tgaatatgtcggagtgtttgaagaaatcttcaccagaagatgataaaaaaggatgtgaac 131
|| || || || || || || || || || || || || || || || || || || || ||
QY 132 atatgataaaaatgtataaagaacagcgtagtaagataaaaagcagcaactlaaacctca 191
|| || || || || || || || || || || || || || || || || || || || ||
Db 1320 tgaatatgtagaagtttgaagaaatcttcaccagaagataaaaatgtaaaaagtcacaatgg 137
|| || || || || || || || || || || || || || || || || || || || ||
QY 192 aattccgaaaagataaatccgaagtgtgcagcttatgtgaattcccgatgtcgtatattaa 251
|| || || || || || || || || || || || || || || || || || || || ||
Db 1380 aattatgtagaagtttgaagaaatcttcaccagaagataaaaatgtaaaaagttbaaatgaaat 143
|| || || || || || || || || || || || || || || || || || || || ||
QY 252 agaacacgatctccagagcacgaacacacctgacataatagagtggtgaagcttgtc 311
|| || || || || || || || || || || || || || || || || || || || ||
Db 1440 agtcaagaagcttgaagaaatcttcaccagaagataaaaatgtaaaaagtcacaatgaataatg 149
|| || || || || || || || || || || || || || || || || || || || ||
QY 312 agaagaanaatgtaacactagatgtatccaataatttcaaattgcagagacacacttcatatga 371
|| || || || || || || || || || || || || || || || || || || || ||
Db 1500 agaggttgaagaataattctaccagaagataaaaatgtaaaaagttcacacatgaaatagtagag 155
|| || || || || || || || || || || || || || || || || || || || ||
QY 372 ccgctcgacactcataatttaccaatctttaagaagccaaaaaaggtagtattgtgtactt 431
|| || || || || || || || || || || || || || || || || || || || ||

Db 1560 agtgaagaattcttcaccgaagatataaaatgaaaaaggtcaacttgaaatagtagagt 16139
QY 432 taaagtgtgtaatgaaacacgctraagtataaatgacaagtataagagatgttgaagctac 491
Db 1620 tgaagaattcttcaccgaagaagatataaaatgaaaaggtcacatgaaatagttagagt 16799
QY 492 agatgtgaggttctgattgaaacaanaagtgaaatlaacacatlaacatlaactactg 551
Db 1680 tgaagaattcttcaccgaagataaaaaatgaaaaaagttcaactgtaaatagttagaggttga 17399
QY 552 tgaattacaaatgaaaagacagcggttgggaaaaacgtlaaatctttgtagctacaga 611
Db 1740 agaattcttcaccagaagataaaatgaaaaaagttcacatgaaatagttagaggttgaaga 17999
QY 612 agt 614
Db 1800 aat 1802

```

RESULT 3
US-08-323-170B-1
Sequence 1, Application US/08323170B
Patent No. 5753772
GENERAL INFORMATION:
APPLICANT: Williams, Kim C.
APPLICANT: Kaslow, David C.
TITLE OF INVENTION: Cloning and Expression of PL
TITLE OF INVENTION: falciiparum Transmission-BLO
NUMBER OF SEQUENCES: 4
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, 8th Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/323,170B
FILING DATE: 13-OCT-1994
CLASSIFICATION: 424
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/010,409
FILING DATE: 29-JAN-1993
ATTORNEY/AGENT INFORMATION:
NAME: Quine, Jonathan A.
REGISTRATION NUMBER: F-41,261
REFERENCE/DOCKET NUMBER: 015280-113100US
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ. ID NO. 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 9636 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 149..9556
US-08-323-170B-1

```

Query Match	8.18;	Score 50.6;	DB 1;	Length 9636;
Best Local Similarity	43.5%;	Pred. No. 0.0014;		
Matches 230; Conservative	0;	Mismatches 229;	Indels 0;	Gaps 0;
Oy	84	tatcgataaattatcttcacgataaagaatgaaagaattgacacatatgataaaaa	143	


```

COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414,057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
US-08-021-601-5

```

```

Query Match          7.8%  Score 48.2;  DB 1;  Length 1368;
Best Local Similarity 44.9%  Pred. No. 0.0032;
Matches 182;  Conservative 0;  Mismatches 223;  Indels 0;  Gaps 0;

```

```

QY 106 aagataaagtgtaagattgacacatatgataaataatgtaaaagacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAACACATTGTAATAAATGAAGTAAGGAGGAGGAGCTGTTAAAAAGAGCAGCA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagcaagcttaaacctcaattccgaagaataatcgaagtgccagctat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAGAAAGTACCATCTGATGTTTAAAGATGTTAAAGCAATTGGAGGAAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 atgaaatccagatgctgatatataaagaacaglatataccaggaaccgaacaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTGTGATGCTGATTTACAAACATATATCTTTAGAACCATTTATCTGAAGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caattaataagagtgtaagctttgcagaagaataatgaatcagatagatcaaatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAAAAATAAAGACATTATTGGAAGATGCTTTATTACATGAACATTATGTTATAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaatgcagagacacactttcaatgaccgctcgaacatacaattacaatcttaagca 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAAAAAAGAGATGTAACCCGCTACTGTATCAATCTTGGAAGATTATGTAAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgctgacttaagttgtaataagaaacgtaaglatataaag 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ACTGAAAGGCACTGAACGTTTATTATGAATAAGTATATATCAAGGATATTTTTA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataagagatgtaagcctcaacgagatgtaagttcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 AGTAAATTAATCAACCATATCAGAAATTTTATAGATGTTTAAAT 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 6
US-08-082-849B-5
Sequence 5, Application US/08082849B
Patent No. 5677274

```

```

GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Ariora, Naveen
APPLICANT: Singh, Yogendra

```

```

APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
OTHER INFORMATION: /product= "L(F(1-254))-TR-DE(401-602)"
US-08-082-849B-5

```

```

Query Match          7.8%  Score 48.2;  DB 1;  Length 1368;
Best Local Similarity 44.9%  Pred. No. 0.0032;
Matches 182;  Conservative 0;  Mismatches 223;  Indels 0;  Gaps 0;

```

```

QY 106 aagataaagtgtaagattgacacatatgataaataatgtaaaagacagcgagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAACACATTGTAATAAATGAAGTAAGGAGGAGGAGCTGTTAAAAAGAGCAGCA 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaaagcagcaagcttaaacctcaattccgaagaataatcgaagtgccagctat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAGAAAGTACCATCTGATGTTTAAAGATGTTAAAGCAATTGGAGGAAG 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 atgaaatccagatgctgatatataaagaacaglatataccaggaaccgaacaacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTGTGATGCTGATTTACAAACATATATCTTTAGAACCATTTATCTGAAGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caattaataagagtgtaagctttgcagaagaataatgaatcagatagatcaaatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAAAAATAAAGACATTATTGGAAGATGCTTTATTACATGAACATTATGTTATAT 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaatgcagagacacactttcaatgaccgctcgaacatacaattacaatcttaagca 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAAAAAAGAGATGTAACCCGCTACTGTATCAATCTTGGAAGATTATGTAAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgctgacttaagttgtaataagaaacgtaaglatataaag 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

DB 421 ACTGAAAAGCGCTGACCTTTATGTAATAGTAAGATATTTATCAAGGATATTTTA 480
Qy 466 acaagataagagatgtaagcctacagatgtaggttcagat 510
DB 481 AGTAATAATTATCAACCATATCAGAAATTTTATGATGTATTAAAT 525

RESULT 7
PCT-US94-01624-5

Sequence 5, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Aroia, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
TITLE OF INVENTION: RELATED METHODS
CORRESPONDENCE ADDRESS:
ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREM
STREET: Steuart Street Tower, 20th Floor, One Market
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 5:
SEQUENCE CHARACTERISTICS:
LENGTH: 1368 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1368
OTHER INFORMATION: /Product=
OTHER INFORMATION: "Lf(1-254)--TR--PE(401-602)"
PCT-US94-01624-5

Query Match 7.8%; Score 48.2; DB 4; Length 1368;

Best Local Similarity 44.9%; Pred. No. 0.0032;

Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

Qy 106 aagaataaagatgaaagattgaacataatgataaaatgtaaaagaaagcgagtaaa 165
DB 121 AAACACATGTGTAATAATAGATAAAAGGAGAGAGCTGTTAAAAAAGACGACAGAGAA 180
Qy 166 gataaagaagcagacgttaacctcaaatccgaaagataaatcgaaagtgccaggtcat 225

DB 181 AACCTACTGAGAAAGTACCATCTGATGTTTAGAGATGTATTAACCAATTGGAGGAAG 240
Qy 226 attgaattccagatgctgatatlaaagaaccagatatccagaccgaacactgaa 285
DB 241 ATATATATTTGTCGATGCTGATATATTACAAAACATATTTCTTTAGAACATTATCGAAGT 300
Qy 286 caattaaatagaggtgtaagccttgcaagaagaatgaactactagatgatacaaatat 345
DB 301 AAGAAAAAATFAAACACATTTTATGGAAAAAGATCTTTTATACATGAACTTATGTATAT 360
Qy 346 tcaatgcygacacacattcatatgaccgtccgaactatcaatitcaatcctaagca 405
DB 361 GCAAAAGAAAGATATGACCCGTACTTGTAATCCAAATCTTCGAGAGATTATGTAGAAAT 420
Qy 406 gccaaaaaagtgatgtggttacttaaaagtgtgaatgaacacgtaagataaaatg 465
DB 421 ACTGAAAAGCGCTGACCTTTTATGTAATAGTAAGATATTTATCAAGGATATTTTA 480
Qy 466 acaagataagagatgtaagcctacagatgtaggttcagat 510
DB 481 AGTAATAATTATCAACCATATCAGAAATTTTATGATGTATTAAAT 525

RESULT 8

US-08-021-601-7
Sequence 7, Application US/08021601
Patent No. 5591631

GENERAL INFORMATION:

APPLICANT: Leppla, Stephen H.
APPLICANT: Kilmpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Aroia, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303

COMPUTER READABLE FORM:

MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514

ATTORNEY/AGENT INFORMATION:

NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880

INFORMATION FOR SEQ ID NO: 7:

SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1416

US-08-021-601-7

Query Match 7.8%; Score 48.2; DB 1; Length 1425;

Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

[illegible]

RESULT 9
US-08-082-849B-7
; Sequence 7, Application US/08082849B

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Alorta, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
TITLE OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor

```

?      LENGTH: 1425 base pairs
?      TYPE: nucleic acid
?      STRANDEDNESS: single
?      TOPOLOGY: linear
?      MOLECULE TYPE: DNA (genomic)
?      HYPOTHETICAL: NO
?      ORIGINAL SOURCE:
?      ORGANISM: Bacillus anthracis
?      FEATURE:
?      NAME/KEY: CDS
?      LOCATION: 1..1416
?      OTHER INFORMATION: /product= "LF(1-254)--Tr--Pe(398-613)"
?
TS-08-082-849E-7

```

Query Match	7.8;	Score 48.2;	DB 1;	Length 1425;
Best Local Similarity	44.9%;	Pred. No. 0.0032;		
Matches 182;	Conservative 0;	Mismatches 223;	Indels 0;	Gaps 0;

QY	106	aagatgaagaatgaagaagcttgcacaaatcgtatgaanaatgttaaaagcagcgagatga	165
Db	130	AAACACATTTGTAAAAATATGAAAGTAAAGGGGACGACACCTGTTAAAAAAGCCACGCGAA	139
QY	166	gataaaagcagcaagcctaaacctcaaatctcgaaagaataatcgaaaglygcagctat	225
Db	190	AAGCTACTTGAAAGAGTACCATCTGATGTTTGTAGAGATGTATTAAGCAATTTGGAGAAAG	249
QY	226	atggaattccagatgctgtatataaagaaccglatatccagagaccagcaacctgaa	285
Db	250	ATATATTTGTGGATGGTGATATTTCAAAACATATATCTTTAGAACGATTATGTGAAGAT	309
QY	286	caatlaaatgaggtgtgaagcttgcagagaanaatgaaacacacagatgaatcaaatatc	345
Db	310	AAGCAAAAAATTAAGACATTTATGCGAAAGATGCTTTATTACATGACACTTATGTTAT	369
QY	346	lcaatgacagacacacttccatgcacgcctcgaaactatcaatctcaaatcaagca	405
Db	370	GCAAAAGAGATATGAAACCCGTAATCTGTAATCCAAATCTTCGGAGATTTATGAGAAAT	429
QY	406	gccaaaaaggtgtagtctgtactcttaaaagttggtaaatgaacaacgtaaagataaagt	465
Db	430	ACTGAAAGGACCTGAAACGTTTATTATGAATATAGTAAAGATATTATTCAGAGGATATTTTA	489
QY	466	acaagatlaagagatgttaagcctacagatgagtgaggatctcaagt	510
Db	490	AGTAAATTTATCAACCATATCCAGAAATTTTATGAGATCTATTAAAT	534

CITY: San Francisco
 STATE: California
 COUNTRY: USA
 ZIP: 94111-3834
 COMPUTER READABLE FORM:
 MEDIUM TYPE: Floppy disk
 COMPUTER: IBM PC compatible
 OPERATING SYSTEM: PC-DOS/MS-DOS
 SOFTWARE: PatentIn Release #1.0, Version #1.30
 CURRENT APPLICATION DATA:
 APPLICATION NUMBER: US/08/082,849B
 FILING DATE: 25-JUN-1993
 CLASSIFICATION: 514
 PRIOR APPLICATION DATA:
 APPLICATION NUMBER: US 08/021,601
 FILING DATE: 12-FEB-1993
 ATTORNEY/AGENT INFORMATION:
 NAME: Weber, Kenneth A.
 REGISTRATION NUMBER: 31,677
 REFERENCE/DOCKET NUMBER: 15280-161-1
 TELECOMMUNICATION INFORMATION:
 TELEPHONE: (415) 576-0200
 TELEFAX: (415) 576-0300
 INFORMATION FOR SEQ. ID NO.: 7:
 SEQUENCE CHARACTERISTICS:

RESULT 10
 PCT-US94-01624-7
 ; Sequence 7, Application PC/TUS9401624
 ; GENERAL INFORMATION:
 ; APPLICANT: Leppla, Stephen H.
 ; APPLICANT: Kilmpel, Kurt R.
 ; APPLICANT: Aroa, Naveen
 ; APPLICANT: Singh, Yogendra
 ; APPLICANT: Nichols, Peter J.
 ; TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
 ; NUMBER OF SEQUENCES: 31
 ; CORRESPONDENCE ADDRESS:
 ; ADDRESSEE: TOWNSEND and TOWNSEND KHOURIE and CREW
 ; STREET: Stewart Street Tower, 20th Floor, One Market
 ; STREET: Plaza
 ; CITY: San Francisco
 ; STATE: CA
 ; COUNTRY: USA
 ; ZIP: 94105
 ; COMPUTER READABLE FORM:
 ; MEDIUM TYPE: Floppy disk.
 ; COMPUTER: IBM PC compatible
 ; OPERATING SYSTEM: PC-DOS/MS-DOS
 ; SOFTWARE: PatentIn Release #1.0, Version #1.25


```

CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 7:
SEQUENCE CHARACTERISTICS:
LENGTH: 1425 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHEICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1416
OTHER INFORMATION: /product=
OTHER INFORMATION: "LF(1-254)--TR--PE(398-613)"
PCT-US94-01624-7

```

```

Query Match      7.8% Score 48.2; DB 4; Length 1425;
Best Local Similarity 44.9%; Pred. No. 0.0032;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aaagataaagatgaagaattgaacaatattgaataatgtaaaagaaagcagagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 130 AACACATTGTAAATAAGAACTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 189
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaagaagcagaactcaatccgaatccgaatccgaatccgaatccgaatccgaat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 190 AAGCTACTTGAAAGAGTACCATCTGATGTTTAAAGATGTTTAAAGCATTTGAGAGAA 249
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 attgaattccagatgctgataatgaagaaccagatatccagaccagcaacacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 230 ATATATATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 309
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caataaataagaggtgaagcttgcagaagaataatgaatccatcagatgaataatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 310 AAGAAATAATTAAGACATTTATGGAAGATGCTTTATTTATCATGAAACATTTATGTA 369
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaattgcagagacacattcattgacgctccgaactatcaattcaaatcttaagaaga 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 370 GCAAAAGAGATATGAACCCCTACTTGTATTCATCTTCGAGAGATATATAGAAAT 429
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgcttaagattgtaaatgaagaacacgtaagataaagtg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 430 ACTGAAGAGCGCTGACCTTTATTTATGAAATAGATATTTATCAAGCATATTTTTA 489
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataaagagtgtaagcctacagatgtaagaggtctcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 490 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATTAAT 534
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 11
US-08-021-601-9
Sequence 9, Application US/08021601
Patent No. 5591631
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Kimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND

```

```

TITLE OF INVENTION: RELATED METHODS
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021,601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
REFERENCE/DOCKET NUMBER: 1414.057
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
US-08-021-601-9

```

```

Query Match      7.8% Score 48.2; DB 1; Length 1524;
Best Local Similarity 44.9%; Pred. No. 0.0033;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

```

QY 106 aaagataaagatgaagaattgaacaatattgaataatgtaaaagaaagcagagtaaa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AACACATTGTAAATAAGAACTAAAGGAGAGAGAGAGAGAGAGAGAGAGAGAGAGAG 180
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 166 gataaagaagcagaactcaatccgaatccgaatccgaatccgaatccgaatccgaat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAAAGAGTACCATCTGATGTTTAAAGATGTTTAAAGCATTTGAGAGAA 240
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 226 attgaattccagatgctgataatgaagaaccagatatccaggaccagcaacacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATATTGTGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 300
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 286 caataaataagaggtgaagcttgcagaagaataatgaatccatcagatgaataatatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGAAATAATTAAGACATTTATGGAAGATGCTTTATTTATCATGAAACATTTATGTA 360
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 346 tcaattgcagagacacattcattgacgctccgaactatcaattcaaatcttaagaaga 405
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 361 GCAAAAGAGATATGAACCCCTACTTGTATTCATCTTCGAGAGATATATAGAAAT 420
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 406 gccaaaaaagtagtattgcttaagattgtaaatgaagaacacgtaagataaagtg 465
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 421 ACTGAAGAGCGCTGACCTTTATTTATGAAATAGATATTTATCAAGCATATTTTTA 480
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
QY 466 acaagataaagagtgtaagcctacagatgtaagaggtctcagat 510
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 481 AGTAAATTAATCAACCATATCAGAAATTTTATGATGATTAAT 525
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```

```

RESULT 12
US-08-082-849B-9

```

```
Sequence 9, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF INVENTION: Related Methods
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESSES:
ADDRESS: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082,849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /product= "Lf(1-254)--TR--PE(362-613)"
US-08-082-849B-9

Query Match 7.8% Score 48.2; DB 1; Length 1524;
Best Local Similarity 44.9%; Pred. No. 0.0033;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

OY 106 aagaataaagatgaaagatgaaacataatgataaataaagaaagcaggcgagataa 165
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 121 AAACACATTGTAATAAGATGAAGTAAGGAGGAGACCTGTTAAAAAGGACAGAA 180
OY 166 gataaaagcagcaagcctaaccataatccgaagaataatcgaagtgagcagctat 225
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 181 AAGCTACTTGAGAAAGATGATGATGATTTAGAGATGATTAAGCATTTGAGGAAG 240
OY 226 atgaaatccagatctgatattaaagaaccagatatatccagcagcaacacctgaa 285
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 241 ATATATTTGTGATGATGATATACAAACATATATCTTTAGAGCATTTATCTGAGAT 300
OY 286 caatlaaatgaggtgtaagcttgcagaaagaataatgaatcactagatgataaatt 345
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
DB 301 AAGCAAAAATTAAGACATTATGGAAGAGATCCTTTATTACATGAACATTATGATAT 360
```

```
RESULT 13
PCT-US94-01624-9
Sequence 9, Application PC/TUS9401624
GENERAL INFORMATION:
APPLICANT: Leppia, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Arora, Naveen
APPLICANT: Singh, Yogendra
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 31
CORRESPONDENCE ADDRESSES:
ADDRESS: TOWNSEND and TOWNSEND KHOURIE and CREW
STREET: Steuart Street Tower, 20th Floor, One Market
STREET: Plaza
CITY: San Francisco
STATE: CA
COUNTRY: USA
ZIP: 94105
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: Patentin Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: PCT/US94/01624
FILING DATE: June 25, 1993
CLASSIFICATION:
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-115
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 543-9600
TELEFAX: (415) 543-5043
INFORMATION FOR SEQ ID NO: 9:
SEQUENCE CHARACTERISTICS:
LENGTH: 1524 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 1..1524
OTHER INFORMATION: /product=
OTHER INFORMATION: "Lf(1-254)--TR--PE(362-613)"
PCT-US94-01624-9

Query Match 7.8% Score 48.2; DB 4; Length 1524;
Best Local Similarity 44.9%; Pred. No. 0.0033;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

OY 106 aagaataaagatgaaagatgaaacataatgataaataaagaaagcaggcgagataa 165
```

```

Db 121 AAACACATTGTAATAATGCAACTAAAGGAGGAGAAAGCTGTTAAAAAGAGCAGCAGAA 180
Oy 166 gataaaagcaagcaagctaaacccaatccgaagaataatcgaaagtgagcagctat 225
Db 161 AAGCTACTTGAGAAAGTACCATCTGATGTTTAGACATGTATTAAGCAATTGGAGAAAG 240
Oy 226 attgaattccagatgctgatatataaagaaccagatataccagaaccgaacacccgtga 285
Db 241 ATATATATTGTTGATGATGATATTAACAAACATATATCTTTAGAGCATTATCTGAAGAT 300
Oy 266 caataaatagaggtgtaagcttgcagagaagaatgtaacactgagatgatacaaatat 345
Db 301 AAGAAATAATTAAGACATTTATGCGAAGATGCTTTATTCATGACATGATATGTATAT 360
Oy 346 tcaattgcagagcacacttcaatlgacgcgccgaactatacttaacaattcaaga 405
Db 361 GCAAAAGAAAGGATATGACACCCCTACTTGTATTCATCTTCGGAAGATTATAGAAAAT 420
Oy 406 gccaaaaaagtagtagtggtagtacttaaaagtggtaaatgaacacgtaagataaaatg 465
Db 421 ACTGAAAGGCGCTGACGTATTTATGAAATAGTAAGATATTTATCAAGGGATATTTTA 480
Oy 466 acaagataagagatgtaagcctacagatgtaagagttcagat 510
Db 481 AGTAAATTAATCAACCATATTCAGAAATTTTATGATTAAT 525

```

RESULT 14
US-08-021-601-1
Sequence 1, Application US/08021601

```

GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Nichols, Peter J.
APPLICANT: Atora, Naveen
TITLE OF INVENTION: ANTHRAX TOXIN FUSION PROTEINS AND
NUMBER OF SEQUENCES: 12
CORRESPONDENCE ADDRESS:
ADDRESSEE: Needle & Rosenberg, P.C.
STREET: 133 Carnegie Way, Suite 400
CITY: Atlanta
STATE: Georgia
COUNTRY: USA
ZIP: 30303
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.25
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/021.601
FILING DATE: 19930212
CLASSIFICATION: 514
ATTORNEY/AGENT INFORMATION:
NAME: Spratt, Gwendolyn D.
REGISTRATION NUMBER: 36,016
TELECOMMUNICATION INFORMATION:
TELEPHONE: 404/688-0770
TELEFAX: 404/688-9880
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3291 base pairs
TYPE: NUCLEIC ACID
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
ORGANISM: Bacillus anthracis

```

```

FEATURE:
NAME/KEY: CDS
LOCATION: 580..2907
US-08-021-601-1

```

Query Match 7.8%; Score 48.2; DB 1; Length 3291;
Best Local Similarity 44.9%; Pred. No. 0.004;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

```

Oy 106 aaagataaagatgaaagattgaacatatgataaaatgtaaaagaacagcgagtaaa 165
Db 700 AAACACATTGTAATAATGCAAGTAAAGGAGGAGAAAGCTGTTAAAAAGAGCAGCAGAA 759
Oy 166 gataaaagcaagcaagctaaacccaatccgaagaataatcgaaagtgagcagctat 225
Db 760 AAGCTACTTGAGAAAGTACCATCTGATGTTTAGACATGTATTAAGCAATTGGAGAAAG 819
Oy 226 attgaattccagatgctgatatataaagaaccagatataccagaaccgaacacccgtga 285
Db 820 ATATATATTGTTGATGATGATATTAACAAACATATATCTTTAGAGCATTATCTGAAGAT 879
Oy 286 caataaatagaggtgtaagcttgcagagaagaatgtaacactgagatgatacaaatat 345
Db 880 AAGAAATAATTAAGACATTTATGCGAAGATGCTTTATTCATGACATGATATGTATAT 939
Oy 346 tcaattgcagagcacacttcaatlgacgcgccgaactatacttaacaattcaaga 405
Db 940 GCAAAAGAAAGGATATGACACCCCTACTTGTATTCATCTTCGGAAGATTATAGAAAAT 999
Oy 406 gccaaaaaagtagtagtggtagtacttaaaagtggtaaatgaacacgtaagataaaatg 465
Db 1000 ACTGAAAGGCGCTGACGTATTTATGAAATAGTAAGATATTTATCAAGGGATATTTTA 1059
Oy 466 acaagataagagatgtaagcctacagatgtaagagttcagat 510
Db 1060 AGTAAATTAATCAACCATATTCAGAAATTTTATGATTAAT 1104

```

```

RESULT 15
US-08-082-849B-1
Sequence 1, Application US/08082849B
Patent No. 5677274
GENERAL INFORMATION:
APPLICANT: Leppla, Stephen H.
APPLICANT: Klimpel, Kurt R.
APPLICANT: Atora, Naveen
APPLICANT: Nichols, Peter J.
TITLE OF INVENTION: Anthrax Toxin Fusion Proteins and
NUMBER OF SEQUENCES: 35
CORRESPONDENCE ADDRESS:
ADDRESSEE: Townsend and Townsend and Crew LLP
STREET: Two Embarcadero Center, Eighth Floor
CITY: San Francisco
STATE: California
COUNTRY: USA
ZIP: 94111-3834
COMPUTER READABLE FORM:
MEDIUM TYPE: Floppy disk
COMPUTER: IBM PC compatible
OPERATING SYSTEM: PC-DOS/MS-DOS
SOFTWARE: PatentIn Release #1.0, Version #1.30
CURRENT APPLICATION DATA:
APPLICATION NUMBER: US/08/082.849B
FILING DATE: 25-JUN-1993
CLASSIFICATION: 514
PRIOR APPLICATION DATA:
APPLICATION NUMBER: US 08/021,601
FILING DATE: 12-FEB-1993
ATTORNEY/AGENT INFORMATION:
NAME: Weber, Kenneth A.

```

```
REGISTRATION NUMBER: 31,677
REFERENCE/DOCKET NUMBER: 15280-161-1
TELECOMMUNICATION INFORMATION:
TELEPHONE: (415) 576-0200
TELEFAX: (415) 576-0300
INFORMATION FOR SEQ ID NO: 1:
SEQUENCE CHARACTERISTICS:
LENGTH: 3291 base pairs
TYPE: nucleic acid
STRANDEDNESS: single
TOPOLOGY: linear
MOLECULE TYPE: DNA (genomic)
HYPOTHETICAL: NO
ANTI-SENSE: NO
ORIGINAL SOURCE:
ORGANISM: Bacillus anthracis
FEATURE:
NAME/KEY: CDS
LOCATION: 580..2907
OTHER INFORMATION: /product= "lethal factor"
US-08-082-849B-1
```

```
Query Match 7.88; Score 48.2; DB 1; Length 3291;
Best Local Similarity 44.98; Pred. No. 0.004;
Matches 182; Conservative 0; Mismatches 223; Indels 0; Gaps 0;

QY 106 aagaataaagatgaaagattgaaacaataatgaaagaagacgagatga 165
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 700 AAGACATTTGTAATAATGAAAGTAAGGAGGAGAGAGCTTTAAATAAGGACAGAA 759

QY 166 gataaagaagcaagctaaacctcaaatccgaagaataaalcgaagtgccagctat 225
    - - - - - | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 760 AACCTACTTGAGAAAGTACCATCTGATGTTTAGAGATGTTAAAGCAATTGGAGAAAG 819

QY 226 attgaattccagatgclgatatlaaagaaccaglatatccagagccagcaaccctgaa 285
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 820 ATATATATTTGTGATGGTATGATTTACAAACATATATCTTTGAAAGCATTTATCTGAAGAT 879

QY 286 caattaatagaagtgtaagcttgcagaagaagaatgaatcactcagatgacaaatat 345
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 880 AAGAAAAAATAAAGACATTTATGGGAAAGATGCTTTATTTACATGAACATTAATATAT 939

QY 346 tcaatgcagagacacacttcaatgaccgctccgaactacaattacaacatctlaaagca 405
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 940 GCAAAAGAGAGATATGAACCGTACTTGTAATCCAATCTTCGAAAGATTATGTGAATAAT 999

QY 406 gccaaaaaagtagtagtgtaacttlaaagttgtagatgaacacgtaagataaagtg 465
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 1000 ACTGAAAAGCAGTGAAGCTTTATTTATGAAATAGTAAGATATTATCAAGGGATATTTTA 1059

QY 466 acaagataagagatgtaagcctacagatgtagagttcagat 510
    ||| . . . | ||| . . . | ||| . . . | ||| . . . | ||| . . . |
DB 1060 AGTAAATTTAATCAACATATCAGAAATTTTATGATGATTAAT 1104
```

Search completed: March 14, 2001, 14:28:04
Job time: 27600 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 06:46:34 ; Search time 1314.13 Seconds
(without alignments)
2418.424 Million cell updates/sec

Title: US-09-292-437-2

Perfect score: 621
Sequence: 1 atgaaaaatgacaaatcg.....tagctacagaagtaataaa 621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 1118133 segs, 2558875100 residues

Total number of hits satisfying chosen parameters: 2236266

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

GenEmbl:*

1: gb_da1:*
2: gb_da2:*
3: gb_om:*
4: gb_ov:*
5: gb_ph:*
6: gb_pi1:*
7: gb_pi2:*
8: gb_pi3:*
9: gb_pi4:*
10: gb_pi5:*
11: gb_pi6:*
12: gb_pi7:*
13: gb_pi8:*
14: gb_pi9:*
15: gb_pi10:*
16: gb_pi11:*
17: gb_pi12:*
18: gb_pi13:*
19: gb_pi14:*
20: gb_pi15:*
21: gb_pi16:*
22: gb_pi17:*
23: gb_pi18:*
24: gb_pi19:*
25: gb_pi20:*
26: gb_pi21:*
27: gb_pi22:*
28: gb_pi23:*
29: gb_pi24:*
30: gb_pi25:*
31: gb_pi26:*
32: gb_pi27:*
33: gb_pi28:*
34: gb_pi29:*
35: gb_pi30:*
36: gb_pi31:*
37: gb_pi32:*
38: gb_pi33:*
39: gb_pi34:*
40: gb_pi35:*
41: gb_pi36:*
42: gb_pi37:*
43: gb_pi38:*

44: em_htg2:*
45: em_htg3:*
46: em_hum5:*
47: gb_pi3:*
48: gb_pi5:*
49: gb_pi8:*
50: gb_pi9:*
51: gb_pi10:*
52: gb_pi11:*
53: gb_pi12:*
54: gb_pi13:*
55: gb_pi14:*
56: gb_pi15:*
57: gb_pi16:*
58: gb_pi17:*
59: gb_pi18:*
60: em_hum4:*
61: em_hum5:*
62: em_hum6:*
63: em_hum7:*
64: em_hum8:*
65: gb_pi18:*
66: gb_pi19:*
67: gb_pi20:*
68: gb_pi21:*
69: gb_pi22:*
70: gb_pi23:*
71: gb_pi24:*
72: gb_pi25:*
73: gb_pi26:*
74: em_hum8:*
75: em_hum9:*
76: em_hum10:*
77: gb_pi6:*
78: gb_pi7:*
79: gb_pi8:*
80: gb_pi9:*
81: gb_pi10:*
82: gb_pi11:*
83: em_hum9:*
84: gb_pi24:*
85: gb_pi28:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Length	DB ID	Description
1	621	100.0	1256 1	AF162687 Staphyloc
2	340.2	54.8	2356 1	AF269739 Staphyloc
3	315.2	50.8	4017 1	AF269386 Staphyloc
4	72.4	11.7	152409 84	PFMALP1
5	68	11.0	196490 29	AC005507
6	67.2	10.8	2069 21	E10125
7	67.2	10.8	3399 21	E10126
8	66.8	10.8	1686 21	E08995
9	66.6	10.7	131271 40	AC015927
10	66.4	10.7	1324 56	PERESAR1
11	66.4	10.7	4591 81	A00661
12	65.4	10.5	7218 31	I66494
13	64.4	10.4	12029 31	AE001373
14	64.2	10.3	201289 49	AC021008
15	64	10.3	879 79	CNS01JRG
16	64	10.3	51014 39	AC011856
17	63	10.1	143585 39	AC013349
18	62.8	10.1	5420 32	AF056936
19	62.4	10.0	202748 84	CNS05TEB
20	62	10.0	83440 51	AC024285
21	62	10.0	153477 29	AC006278

C	22	62	10.0	166050	51	AC024352	AC024352 Homo sapi
C	23	61.2	9.9	7852	31	AB018798	AB018798 Plasmodiu
C	24	60.6	9.8	11829	31	AE001376	AE001376 Plasmodiu
C	25	60.6	9.8	192581	84	PFMAL13P1	AL049180 Plasmodiu
C	26	60.4	9.7	14760	31	AE001410	AE001410 Plasmodiu
C	27	60.2	9.7	156550	40	AC015830	AC015830 Homo sapi
C	28	60	9.7	825	32	AF206632	AF206632 Plasmodiu
C	29	59.8	9.6	2349	81	A00764	A00764 Synthetic P
C	30	59.8	9.6	4739	56	PFAGLURPA	M59706 P. falciparu
C	31	59.8	9.6	192929	29	AC005505	AC005505 Plasmodiu
C	32	59.4	9.6	153936	65	AC079863	AC079863 Homo sapi
C	33	59.2	9.5	170132	84	CNS057DS	AL357093 Homo sapi
C	34	59	9.5	840	7	CNS0180K	AL110675 Botrytis
C	35	59	9.5	74119	54	AC036177	AC036177 Homo sapi
C	36	59	9.5	113880	56	PFMAL13P4	AL008970 Plasmodiu
C	37	58.8	9.5	175957	40	AC015961	AC015961 Homo sapi
C	38	58.6	9.4	88932	50	AC022680	AC022680 Homo sapi
C	39	58.4	9.4	110000	84	PFMAL13P2_1	Continuation (2 of
C	40	58.2	9.4	95764	40	AC015732	AC015732 Homo sapi
C	41	58.2	9.4	200125	54	AC034105	AC034105 Homo sapi
C	42	57.8	9.3	81120	50	AC022851	AC022851 Homo sapi
C	43	57.8	9.3	164399	56	PFMAL13P6	298551 Plasmodium
C	44	57.8	9.3	245802	29	AC006279	AC006279 Plasmodiu
C	45	57.4	9.2	3085	56	DD060169	U60169 Dicyostell

ALIGNMENTS

RESULT 1

LOCUS	AF162687	1256 bp	DNA	BCT	11-AUG-1999
DEFINITION	Staphylococcus aureus sortase (srta) gene, complete cds.				
ACCESSION	AF162687				
VERSION	AF162687.1	GI:5726435			
KEYWORDS	Staphylococcus aureus.				
SOURCE	Staphylococcus aureus.				
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.				
REFERENCE	1 (bases 1 to 1256)				
AUTHORS	Mazmanian, S.K., Liu, G., Ton-That, H. and Schneewind, O.				
TITLE	Staphylococcus aureus sortase, an enzyme that anchors surface proteins to the cell wall				
JOURNAL	Science 285 (5428), 760-763 (1999)				
MEDLINE	99357874				
PUBMED	10427003				
REFERENCE	2 (bases 1 to 1256)				
AUTHORS	Mazmanian, S.K., Liu, G., Ton-That, H. and Schneewind, O.				
TITLE	Direct Submission				
JOURNAL	Submitted (24-JUN-1999) Microbiology and Immunology, UCLA, 10833 E				
Conte Avenue, Los Angeles, CA 90095, USA					
FEATURES	Location/Qualifiers				
source	1..1256				
	/organism="Staphylococcus aureus"				
	/strain="8325-4"				
gene	/db_xref="taxon:1280"				
	483..1103				
	/gene="srta"				
CDS	483..1103				
	/gene="srta"				
	/note="transpeptidase"				
	/codon_start=1				
	/transl_table=11				
	/product="sortase"				
	/protein_id="A008437.1"				
	/db_xref="GI:5726436"				
	/translation="MKWTNRMTAGVLLVLAAYLFAKPHIDNYLHDKDKETIO				
	YDKNVKQASDKQQAQPIQPKDKSVAGYIEIPDAIKKEPYGPGATGEOLNRGVS				
	FAEENSIDQONISIAIGHFIDRPNOFTULKAAGKSMVFPVGNTRKYMKSIND				
	VKPTDYGIVDEQKQKQKQLTLITCDYNEKTVGWEKKRIVATIEVK"				

BASE COUNT

470 a 185 c 234 g 367 t

Query Match 100.0%; Score 621; DB 1; Length 1256;
Best Local Similarity 100.0%; Pred. No. 6, 7e-89;
Matches 621; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

Qy	1	atgaaaaaatgacaaatcgatlaaagcaatcgctgctgcttaccctagtgca	60		
Db	483	ATGAAAATATGACAAATCGATTATGCAATCGCTGCTGCTTATCTTACGTCGA	542		
Qy	61	gcatattgttctgctaaacccatatactatatacttccacgataaagaatgaa	120		
Db	543	GCATATTTGTTTGTCTAAACGACATATCGATTATCTTCAAGATAAAGATATGAA	602		
Qy	121	aagatttgacataatgataaataatgataaagacagagatgataaagaagca	180		
Db	603	AAGATTGACATATGATATAAATGTAAAGACAGCGAGTAAAGATATAAGACGCA	662		
Qy	181	gctaaacctcaaatccgaaagataaalcgaagtgagcgtatattgaaattccagat	240		
Db	663	GCTAAACCTCAAAATTCGAAAGATTAATCGAAAGTGCAGGCTATATTGAATTCAGAT	722		
Qy	241	gctgatatataaagaccgatatatccagagaccgaacacgtgaaacattaatagagt	300		
Db	723	GCTGATATTAAGAACCGATATATCCAGACGACACACCTGAAACATTTAATAGAGCT	782		
Qy	301	gtaagcttgacagaagaatgaaatcactagatgataaataattcaattgacagacac	360		
Db	783	GTAAGCTTTGACAGAAGAATAATGATCACTAGATGATCAAAATATTTCATTCAGACAC	842		
Qy	361	acttcattgacgctcgcgaactatcaattcaaatctttaaagcagcacaagaagtagt	420		
Db	843	ACTTTCATTGACCGCCGCAACTATCAATTTCAAAATCTTAAAGCCGCAAAAAGTAGT	902		
Qy	421	atggtgatacttaaaagttggtgaatgaacacgtaagatataaataagacatagagat	480		
Db	903	ATGGTACTTTAAAGTTGATGTAATGAACACGTAGTAAATATCAAGATATAAGAGAT	962		
Qy	481	gttaagcttacagatgtgagagcttcagatgataaagaagtaagaataaataa	540		
Db	963	GTTAAGCTTACAGATGTAGAGCTTACATGATGACAAAAAGTAAAGATATAAACA	1022		
Qy	541	ttaattacttgatgatacaatgaagaaagacagcgcttgaggaaaacgataaattc	600		
Db	1023	TTAATTACTTGTGATGATTAATGAATAAAGACAGCGCTTGGAATAAAGCTAATCTTT	1082		
Qy	601	gtagctacagaatgcaataa 621			
Db	1083	GTAGCTACAGAGTCAATAATA 1103			
RESULT 2					
LOCUS	AF269739	2356 bp	DNA	BCT	01-AUG-2000
DEFINITION	Staphylococcus epidermidis strain SRI clone step.1020h12 genomic				
ACCESSION	AF269739				
VERSION	AF269739				
KEYWORDS	Staphylococcus epidermidis.				
SOURCE	Staphylococcus epidermidis.				
ORGANISM	Bacteria; Firmicutes; Bacillus/Clostridium group; Staphylococcus.				
REFERENCE	1 (bases 1 to 2356)				
AUTHORS	Kimmerly, W.D., Taylor, J. David, Nelson, A.J., Godlewski, M.M., Rubino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Listenebee, S., Ashanti, C., Altschuler, G., Memo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S., and Furdon, P. J.				
TITLE	Transposon-mediated sequencing of the Staphylococcus epidermidis				
JOURNAL	Unpublished				
REFERENCE	2 (bases 1 to 2356)				

AUTHORS	Taylor, J. David, Kimerly, W.J., Nelsen, A.J., Godlewski, M.M., Rudino, M.A., Nelson, F.J., Rivers, P.R., Torruella-Miller, I., Lustendeb, S., Ashanti, C., Alshuller, G., Memo, L., Shepherd, N.S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D.H., Miller, G.S. and Furdon, P. J.
TITLE	Direct Submission
JOURNAL	Submitted (22-May-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3398, USA
FEATURES	location/Qualifiers
source	1. .2356 /organism="Staphylococcus epidermidis" /strain="SRI" /db_xref="taxon:1282" /clone="step.1020h12"
BASE COUNT	777 a 371 c 332 g 876 t
ORIGIN	
Query Match	54.8%, Score 340.2; DB 1; Length 2356;
Best Local Similarity	73.1%; Pred. No. 7.9e-45;
Matches 454; Conservative 0; Mismatches 158; Indels 9; Gaps 1;	
OY 1	atgaaaaaatggacaacatcgattaatgcgaatcgcgtgtgtgtacttatcttagtgcga 60
Db 1222	ATGAAGCGACGTGATTAATAGATTAAATCACTTAAATAGCCGATTTCTTATCATTTTACTC 1163
OY 61	gcaatactgtttgttcgaaccacatbcgatcatatbtcttcacgaataagaataagtgaa 120
Db 1162	ATTATTATTATTTCCTCAAACACCATATTCATATATTATTCATCAATGAATAAAGATTACGATCAT 1103
OY 121	aagattgcacaatatgatataaaaatgtlaaaagacagcgagtgtaaagtataaaaagacgaa 180
Db 1102	AAAAATTGAAAATTATGATATAAAGAAGAAAAGACAAGCAAAGACATCTNA-----AA 1052
OY 181	gctaaacctcaaatccgaaagaataaatcogaagtgtgcaggctatatgtgaaattccagat 240
Db 1051	TCGAGCGCCAAGATACCTCTCCGATTAATCTAAATGGCTGCTTATATGAACTTCACGAT 992
OY 241	gctgatlattaagaaccagatlatccaggaccagacaacactgaaacaatltaaatagagt 300
Db 991	GCACAATTAANAAGACACCGATATACCTCTGTCAGCAACCCAGAACCACTCAATAGAGGT 932
OY 301	gtaagcttctgcagaagaataatgaatcactagatgatacaaaatattcaattcagagcac 360
Db 931	GTTAGTTTTTGCAAGAAGTGACGAATCTCTTATATCAACGAANATTTCATTCTGCTGTCAT 872
OY 361	acttcatltgacgcgtccgaactataccaatttcaaacctttaaagcgccaaaagtgatg 420
Db 871	ACGTTTACAGATTCGCTCCACACTATCAATTTCAAAATTTAAATCACGCCAAAATCGTGAGT 812
OY 421	atgggtgaactttaaagtttgtgtgaatgaagaacagtgtaagtblaaaabgacaaugataaagaat 480
Db 811	AAAGGTATTTTAAACTGTGAATATAAATCTGAAGATGTAAATATACCTTAAATACGTGAT 752
OY 481	gttaagcctacaagatgtatgaggttctagatgcaacaataaagtgaaagataaacaattaaca 540
Db 751	GTTAAGCCTACGAGAGTTAAGGTATTATNACGACACATCTCTAATTAAGAAAAATCAATTAA 692
OY 541	ttaattactctgtagatgattacaatgtaaaagacagcgctgtttgggaaaaacytaaatcttt 600
Db 691	TTTAATTACTTCGATCGACTATTAACGAAGAAACGCGTGTGTTGGCAAAACGAAGAAATATTTC 632
OY 601	gtagctacagaagatcaataaa 621
Db 631	ATAGCTACACAATAATTAACCTAA 611
RESULT 3	AF269386 4017 bp DNA BCT 03-Aug-2000
LOCUS AF269386/c	Staphylococcus epidermidis strain SRI clone step.1003a02 genomic sequence.

ACCESSION	AF269386
VERSION	AF269386.1
KEYWORDS	GI:9664707
SOURCE	
ORGANISM	Staphylococcus epidermidis. Staphylococcus epidermidis Bacteria; Firmicutes; Bacillus/Clostridium group; Bacillus/Staphylococcus group; Staphylococcus.
REFERENCE	1 (bases 1 to 4017) Kimmerly, W.J., Taylor, J. David, Nelsen, A. J., Godlevski, M. M., Rubino, M. A., Nelsen, F. J., Rivers, P. R., Torturella-Miller, I., Listeenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Purdon, P. J.
TITLE	Transposon-mediated sequencing of the Staphylococcus epidermidis genome
JOURNAL	Unpublished
REFERENCE	2 (bases 1 to 4017) Taylor, J. David, Kimmerly, W.J., Nelsen, A.J., Godlevski, M. M., Rubino, M. A., Nelsen, F. J., Rivers, P. R., Torturella-Miller, I., Listeenbee, S., Ashanti, C., Altschuller, G., Mamo, L., Shepherd, N. S., Fuchs, R., Fleming, T., Guan, X., Du, L., Cain, D. H., Miller, G. S. and Purdon, P. J.
AUTHORS	Direct Submission Submitted (19-MAY-2000) Departments of Genomic Sciences and Bioinformatics, Genetics Directorate, Glaxo Wellcome, Inc., 5 Moore Drive, Research Triangle Park, North Carolina 27709-3390, USA
TITLE	Location/Qualifiers
FEATURES	1. 4017 /organism="Staphylococcus epidermidis" /strain="SR1" /db_xref="taxon:1282" /clone="step_1003a02"
SOURCE	
BASE COUNT	1270 a 647 c 616 g 1484 t
ORIGIN	

Query Match	50.88	Score 315.2	DB 1	Length 4017
Best Local Similarity	75.0%	Pred. No. 6.1e-41		
Matches 411	Conservative 0	Mismatches 128	Indels 9	Gaps 1
Qy 74	ctaaaccacatcatgatataatctatcttcacagtaaaagataaagatgtaaaagattacaat	133		
Dy 4016	CAAGCCATATATGATGATATATTATCTACTAGTGAAGAAAGATACGATCATTAATTTGAAATT	3957		
Qy 134	atgataaaatgtaaaagaacagcgagtaaaagataaaagcgaagctaaacctcaa	193		
Dy 3956	ATGATTAATAAACGAAAGAAAGACAGCAAAAGACATCTAA-----TCGAGCCCAAGA	3906		
Qy 194	ttccgaagaagataatcgaaaglygcagctatattgaaattccagatgctatataag	253		
Dy 3905	TACCTTCGGATTAATCTTAAATGGTGCTTTATATGAAAGTTCCAGATGCCAATAATAAAG	3846		
Qy 254	aaccagatataccagagcagcaaaccttgacaataaataagagtgtaagcttgcag	313		
Dy 3845	AACCATATACCTTCGTGCCACACACCCAGACAACATCAATATAGAGTGTGTTTGCGAG	3786		
Qy 314	aagaaatgatactagatgatcaaaatatttcaatltgcagacacattlctgacc	373		
Dy 3785	AAGGACGAAATCTCTTATATCAACAGATATTTCATTCGTGCATATCGTTACAGATC	3726		
Qy 374	gtccgaactatcaatttcaaaactttaaagcagccaaagagtagtgglyacttta	433		
Dy 3725	GTTGCGACATATTCATTTTCAAAATTTTAAATATGCGCAAAATCGTAGTAAAGTGATTTT	3666		
Qy 434	aagttgtaatgaaacagtaagataaaatgaaacagataagagatgtaagctcag	493		
Dy 3665	AAACTGGAAATCAACATCTGAAAGATTAATAATTAACATAATTCGTATGTTAAGCTTAAG	3606		
Qy 494	atgtagaggtctcatgatgatacaaaaaglytaagaatacaataaacatttaacttctg	553		
Dy 3605	AGGTTAAGGTATTAGACCAACATCCTATATAGGAAAAATCATTTAATATTACTTCGC	3546		
Qy 554	atgatatacgaataagacagcgcttggggaataaacgtlaaaactcttgtagctacgaag	613		

Db	44675	AATTAAGCGTAAAGATACCAAAATATTAAATATCCAAAGATATTATTAATAATGAAAAGTATT	44616
Qy	244	gataltaagaagcaagatataccaggaccagcaacacctgaacaaatlaaagagtgta	303
Db	44615	AATATATATGATTAATATTAATAATTCCAGACAGAAATTAAGATATATACAGAAATGAAA	44556
Qy	304	agccttgcagaagaaatggaatcactagaatgatacaaaatlctcaatgtccaggacacact	363
Db	44555	AGTGCAGGTATTAAGAAAGAAATGTGGAAATATGATATATTCATTAATTTATTAATTAAT	44496
Qy	364	ttcatctgacgcgtccgaactatcaatcttcaaaccttaagaagccaaaaggtagatg	423
Db	44495	TTTCATATGTAAGAAAGATATCTGTAGATGCTATTATTGTTCGTGAGAAAGACAAAGACTGAACCG	44436
Qy	424	gtgtaactttaaagtgtgtaacgtgaacacgylaaglaataaatgacaagatataagagatgt	483
Db	44435	ATGAAAGGGAATGATTAATATATGTGAAAGATCAAAAAGAAAGAAACATGAGAGATTAAT	44376
Qy	484	aagcctaagaatgttagagtgctctgatagaatacaaaaaggtlaaagaataaacatlaacatla	543
Db	44375	AAGAAAGAAAGTATTAAATTAGCAAAAACAGAAAGAGGTGATACATCAAAAAGAAACAAAAA	44316
Qy	544	attactgtgatagtatcaatcaagaaa 569	
Db	44315	AATATGTGCGATTAATAAAATAATAAAAA 44230	
RESULT	6		
ID	E10125	standard; RNA; UNC; 2069 BP.	
XX	AC	E10125:	
XX	SV	E10125.1	
DT	08-OCT-1997	(Rel. 52, Created)	
DT	02-SEP-2000	(Rel. 65, Last updated, Version 2)	
XX	DE	DNA encoding an immunogenicity protein.	
XX	KW	JP 1995284392-A/1.	
XX	OS	unidentified	
OC	OC	unclassified.	
XX	XX		
RN	[1]		
RP	1-2069		
RA	Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;		
RT	"GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND GENE		
RT	RECOMBINANT VACCINE AGAINST CHICKEN LEUCOCYTOZOON";		
RL	Patent number JP1995284392-A/1, 31-OCT-1995.		
RL	DOUBUTSUYOU SEIBUTSUGAKUTEKI SEIZAI KYOKAI, KITASATO INST.:THE.		
XX	XX		
CC	OS	Leucocytozoon caulleryi	
CC	PN	JP 1995284392-A/1	
CC	PD	31-OCT-1995	
CC	PI	19-APR-1994 JP 1994080643	
CC	CC	DOI HIROHITO, NAGAKUCHI YOSHIO, TANAKA YOSHIO, FUJISAKI YUJIRO	
CC	PC	C12M15/09,A61K39/015,C12P21/02;	
CC	CC	strandedness: Double;	
CC	CC	topology: Linear;	
CC	CC	key	
CC	CC	location/Qualifiers	
CC	FT	source	1..2069
CC	FT	organism="leucocytozoon caulleryi"	
CC	FT	strain="shizuoka"	
CC	FT	misc_feature	1..25
CC	FT	note="EcoRI adaptor"	
CC	FT	misc_feature	2045..2069
CC	FT	note="EcoRI adaptor"	
CC	FT	misc_feature	25..2046
CC	FT	product="immunogenicity protein"	
XX	XX		

Query Match	Best Local Similarity	10.8%; Score 67.2; DB 21; Length 2069;
Matches 237; Conservative 0; Mismatches 283; Indels 0; Gaps 0;		
QY 101	acgataaagataaagatgaaagattgacaataatgataaataatgtaaagaacagcgca	160
Db 920	AAGAAAAAGAGAGAGTACACATGAAAGAAATGAAAAAGAGAGCATGAGAGTAATAC	979
QY 161	gtaaaagataaaagagagaagcctaactcaattccgaaagataatcgaagtcgcaag	220
Db 980	ATGAAGAAGAAAAACAAGACATGACATGAAAGAAAAAGAAAAAGACATGTAAGAG	1039
QY 221	gcatattgaaattccagatgctgataltaagaacacagatatccagagaccagaacac	280
Db 1040	TAATTCATGAGAGAAAAAGAGAGATACCATGAAAGAACAGACAGAAAAAGTAACAC	1099
QY 281	ctgaacaaatlaaatagaggtgtaagctttgcagaaataatgtaatacactagatgcaaa	340
Db 1100	ATGAAGAAGAGAGAGAAAAAGTAATACATGAAAGAGAAAAAGAGACAGTAATACATGAAG	1159
QY 341	ataattcaattgcagagacacacttcaattgacgcgtccgaactataatcaattacaactta	400
Db 1160	AAGAAAAAGAGAACTAACACATGAAAGAGAAAAAGAGAGATACACATGAGAGAGAA	1219
QY 401	aagcagcccaaaaaggtgtagtactgtaagttgtaattgaataaacaagtaagtata	460
Db 1220	AAGAGAGAGTAAACACATGAGAGAAAAAGAGAGTAAACATGAGAGAGAAAAAGAGAG	1279
QY 461	aaatgacagatlaaagagatgttgaagcctacagatgtagagttctagatgacacaaaag	520
Db 1280	AAGTAAACATGAGAGAGAGAAAAAGTAACATGAAAGAGAAAAAGAGAGTAATAC	1339
QY 521	gtaaaagataaacaatlaacatlaacttacttgtagatgattacaatgaaagacagcgctt	580
Db 1340	ATGAAGAAGAGAAAAACAAGATACCATGAAAGAGAGAAAAAGTAACACATGAAGAG	1399
QY 581	gggaaaaacgtaaaaactcttgtagtcaagaagttcaata 620	
Db 1400	AAGAAAAAGTAAACATGAGAGAGAGAAAAAGTAACATA 1439	
RESULT 7		
ID E10126	standard; DNA; UNC; 3399 BP.	
XX AC E10126;		
XX SV E10126.1		
DT 08-OCT-1997 (Rel. 52, Created)		
DT 02-SEP-2000 (Rel. 65, Last updated, Version 2)		
DE DNA encoding an immunogenicity protein of Leucocytozoon caulleryi fused to		
DE maltose-binding protein.		
XX JP 1995284392-A/2.		
XX unclassified		
XX OS unclassified		
XX CC unclassified		
FN [1]		
RP 1-3399		
FP Doi H., Nagakuchi Y., Tanaka Y., Fujisaki Y.;		
FT "GENE CLONE MANIFESTING CHICKEN LEUCOCYTOZOON IMMUNOGENIC PROTEIN AND GENE		
FT /db_xref="taxon:32644"		
FT /organism="unidentified"		
XX Sequence 2069 BP; 1216 A; 167 C; 443 G; 243 T; 0 other;		


```

* 2766 27865: gap of 100 bp
* 2786 28807: contig of 942 bp in length
* 28808 28907: gap of 100 bp
* 28908 29847: contig of 940 bp in length
* 29848 29947: gap of 100 bp
* 29948 30897: contig of 950 bp in length
* 30898 30997: gap of 100 bp
* 30998 31912: contig of 915 bp in length
* 31913 32012: gap of 100 bp
* 32013 32917: contig of 905 bp in length
* 32918 33017: gap of 100 bp
* 33018 33967: contig of 950 bp in length
* 33968 34067: gap of 100 bp
* 34068 35010: contig of 943 bp in length
* 35011 35110: gap of 100 bp
* 35111 36047: contig of 937 bp in length
* 36048 36147: gap of 100 bp
* 36148 37083: contig of 936 bp in length
* 37084 37183: gap of 100 bp
* 37184 38131: contig of 948 bp in length
* 38132 38231: gap of 100 bp
* 38232 39180: contig of 949 bp in length
* 39181 39280: gap of 100 bp
* 39281 40231: contig of 951 bp in length
* 40232 40331: gap of 100 bp
* 40332 41226: contig of 895 bp in length
* 41227 41326: gap of 100 bp
* 41327 42275: contig of 949 bp in length
* 42276 42375: gap of 100 bp
* 42376 43301: contig of 926 bp in length
* 43302 43401: gap of 100 bp
* 43402 44336: contig of 935 bp in length
* 44337 44436: gap of 100 bp
* 44437 45355: contig of 919 bp in length
* 45356 45455: gap of 100 bp
* 45456 46372: contig of 917 bp in length
* 46373 46472: gap of 100 bp
* 46473 47424: contig of 952 bp in length
* 47425 47524: gap of 100 bp
* 47525 48460: contig of 936 bp in length
* 48461 48560: gap of 100 bp
* 48561 49507: contig of 947 bp in length
* 49508 49607: gap of 100 bp
* 49608 50571: contig of 964 bp in length
* 50572 50671: gap of 100 bp
* 50672 51589: contig of 918 bp in length
* 51590 51689: gap of 100 bp
* 51690 52631: contig of 942 bp in length
* 52632 52731: gap of 100 bp
* 52732 53692: contig of 961 bp in length
* 53693 53792: gap of 100 bp
* 53793 54720: contig of 928 bp in length
* 54721 54820: gap of 100 bp
* 54821 55741: contig of 921 bp in length
* 55742 55841: gap of 100 bp
* 55842 56781: contig of 940 bp in length
* 56782 56881: gap of 100 bp
* 56882 57794: contig of 913 bp in length
* 57795 57894: gap of 100 bp
* 57895 58846: contig of 952 bp in length
* 58847 58946: gap of 100 bp
* 58947 59869: contig of 923 bp in length
* 59870 59969: gap of 100 bp
* 59970 60880: contig of 911 bp in length
* 60881 60980: gap of 100 bp
* 60981 61900: contig of 920 bp in length
* 61901 62000: gap of 100 bp
* 62001 62918: contig of 918 bp in length
* 62919 63018: gap of 100 bp
* 63019 63968: contig of 950 bp in length
* 63969 64068: gap of 100 bp
* 64069 64986: contig of 918 bp in length
* 64987 65086: gap of 100 bp

```

```

* 65087 66019: contig of 933 bp in length
* 66020 66119: gap of 100 bp
* 66120 67033: contig of 914 bp in length
* 67034 67133: gap of 100 bp
* 67134 68116: contig of 983 bp in length
* 68117 68216: gap of 100 bp
* 68217 69171: contig of 955 bp in length
* 69172 69271: gap of 100 bp
* 69272 70184: contig of 913 bp in length
* 70185 70284: gap of 100 bp
* 70285 71229: contig of 945 bp in length
* 71230 71329: gap of 100 bp
* 71330 72293: contig of 964 bp in length
* 72294 72393: gap of 100 bp
* 72394 73305: contig of 912 bp in length
* 73306 73405: gap of 100 bp
* 73406 74335: contig of 930 bp in length
* 74336 74435: gap of 100 bp

```

Query Match 10.7%; Score 66.6; DB 40; Length 131271;
 Best Local Similarity 36.4%; Pred. No. 0.036;
 Matches 198; Conservative 0; Mismatches 346; Indels 0; Gaps 0;

```

Qy 77 aaccacatcgcataatcattcaccgataagaataagaatgaagaattgaacatg 136
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55664 AANNATATATAGAGATGATTAATTATAAAAAAAAAAAAAAAAAAAAAA 55605

Qy 137 ataaaatgtaagaagcagcgagtaagaataagaacagcagctaacctc 196
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55664 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55545

Qy 197 cgaagaataatcgaagctgcagcgtataltgaattccagatgctatlaagac 256
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55544 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55485

Qy 257 cagatataccgcgcgcgcgcacacccgtgaacattgaatagcgttaagcag 316
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55484 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55425

Qy 317 aaatgtaactcagatgatcacaatatcattgcgcgcacattcgtgcgctc 376
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55424 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55305

Qy 377 cgaactatcaatttacaatttaagcagccaaaagtagtattgtaacttaag 436
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55364 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55305

Qy 437 ttggtatgaacacgtaagataataatgcaagataagatgtaagcctaacag 496
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55304 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55245

Qy 497 taggagttcagatcgaacaaagtaagaataacaattgaattacttgctgtg 556
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55244 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 55185

Qy 557 attcaatgaaagcagcgcttgggaaaacgtaaatcttltgactcagaagta 616
    ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 55184 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATANTTTTANAAAAAA 55125

Qy 617 ata 620
    |||
Db 55124 CAAA 55121

```

RESULT 10
 PFRESARI LOCUS 1324 bp mRNA INV 16-FEB-1998
 DEFINITION P.falciparum FC27 Ag46 RESA mRNA for ring-infected erythrocyte
 surface antigen.
 ACCESSION X05182
 VERSION X05182.1 GI:9962
 KEYWORDS antigen; ring-infected erythrocyte surface antigen; surface
 antigen; tandem repeat.

SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
AUTHORS 1 (bases 1 to 1324)
TITLE Commam,A.F., Coppel,R.L., Saint,R.B., Favaloro,J., Crewther,P.E.,
Stahl,H.D., Bianco,A.E., Brown,G.V., Anders,R.F. and Kemp,D.J.
The ring-infected erythrocyte surface antigen (RESA) polypeptide of
Plasmodium falciparum contains two separate blocks of tandem
repeats encoding antigenic epitopes that are naturally immunogenic
in man
JOURNAL Mol. Biol. Med. 2 (3), 207-221 (1984)
MEDLINE 85295381
COMMENT Data kindly reviewed (10-JUL-1987) by Favaloro J.
For FC27 RESA genomic sequence see X05181; for related NF7 CDNA
sequence see X05181.
FEATURES
source
location/Qualifiers
1..1324
/organism="Plasmodium falciparum"
/strain="FC27 (Papua New Guinea isolate)"
/db_xref="taxon:5833"
/clone="FC27 Ag46"
1..955
/gene="RESA"
/gene="RESA"
/codon_start=2
/product="ring-infected erythrocyte surface antigen"
/protein_id="CAA0077.1"
/db_xref="GI:1335718"
/db_xref="SWISS-PROT:P13830"
/translation="NSTYFNENINSVNDNGNSKNISDLSYTDKEILEKIVSYIV
ISLYDIENALNAEQLSDNSVDEKTLKRAOSLKLSIMERYAGGRKNDKKNF
DTKIVGYIMHGISITINTENKNONENPEVHOHNAEENVEHDAEENVEHDAEENVEH
AEENVEHDAEENVEHDAEENVEENVEENVEENVEENVEENVEENVEENVEENVEEN
VEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVE
VEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVE
388..954
repeat_region
BASE COUNT 594 a 85 c 230 g 415 t
ORIGIN
Query Match 10.7%; Score 66.4; DB 56; Length 1324;
Best Local Similarity 47.9%; Pred. No.0.093;
Matches 224; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
OY 81 aacatcgataattcttcacgataaagataagatgaagaattgaacaatatgataa 140
DB 499 ACATGATGCTGAAGAAATGTAGAACATGATGCTGAAGAAATGTAGAAATGTTGA 558
OY 141 aatgtaaaagacagcgagtaagataaaagcagcaagctaaaccctcaatccgaa 200
DB 559 AGAAGTAGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTTGACA 618
OY 201 agataaatcgaagtggcagctataltgaa--altccagatgctgataltaagaacc 257
DB 619 AGTAGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTTGAAGAAAA 678
OY 258 agtatatccagagccagcaaccctgaacatlaaataagtgtaagcttgccagaaga 317
DB 679 TGTGAAGAAATGTTGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTTGAAGAAAA 738
OY 318 aaatgaatcaatagatgatalcaaatataltcgaatgcaagcacttcaatgcaagctgc 377
DB 739 TGATGAGAGAAATGTTGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTTGAAGAAAA 798
OY 378 gaataatcaatltacaaatcctaagaagccaaagaaggtaglatggttactttaagt 437
DB 799 TCAGAGAAATGTAGAAAGAAATGTTGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTAGA 858
OY 438 tggtaataaacaacgtlaaglatataaaatgcaagttataagaagtggttaagcccaagatgt 497
DB 859 ACAAAATGTAGAAAGAAATGTTGAAGAAATGTAGAAAGAAATGTAGAAAGAAATGTATGATGA 918

OY 498 aggaagttcagatgacacaaaagtgtaagataaacaattacaattaat 545
DB 919 AGAAATGTTGAAGAAACACATGAAAGAAATATGATGAATAAAAAAAT 966
RESULT 11
LOCUS A00661 4591 bp DNA PAT 09-JUL-1993
DEFINITION P.falciparum RESA gene for ring-infected erythrocyte surface
antigen.
ACCESSION A00661
VERSION A00661.1 GI:410747
KEYWORDS
SOURCE malaria parasite P. falciparum.
ORGANISM Plasmodium falciparum
REFERENCE 1 (bases 1 to 4591)
AUTHORS Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium.
TITLE ANTIGENS OF PLASMODIUM FALCIPARUM
JOURNAL Patent: WO 8601802-A 1 27-MAR-1986;
FEATURES
source
location/Qualifiers
1..4591
/organism="Plasmodium falciparum"
/db_xref="taxon:5833"
801..4225
/gene="RESA"
/gene="RESA"
/join(801..995,1199..4225)
/codon_start=1
/product="ring-infected erythrocyte surface antigen"
/protein_id="CAA0077.1"
/db_xref="GI:410748"
/db_xref="SWISS-PROT:P13830"
/translation="MRPHAYSMIFSQOYMGKNYKKNKNTYSPDDEPKRMENKSF
KYLCKRGVLPITGLIYLINLNCLYNGSSSGVPTDRCSNLKLTLPVPYADSE
NEIVSÖVFGLEPFKEPFTLESPPIDITNLLGNEKEFTVDNRRYRSNNVEAIPHS
FENPLIVKVLFDYNEKYDNLSRSGDIIKKMOTLMDLMDINKRKYLSLKEKQTY
SQYKQYDMPKFAVESKMTQCILIDOGGENLEERLNSQFMWRYKRYLNEEYRLT
VINOYAMKALSNOIOYSCRKIMNSDISFRKHNELSLERAKAAAEEMKRAQPK
KKSRGWLCCGGGDLITVEPQOEPPYOVVOOVNENGLIPSLASTYNSAIRAYD
TYKGVYLDHETSALTYDEDLILPLEKOKYMDLDTSEESVEENEBEHYDDHVE
EHTADDEHVEEPTVADDEHVEEPTVADDEHVEEPTVADDEHVEEPTVADDEHVEEPTVAD
QTEAAPPTEIETPDLYDILGVANADNMEITERFKLAENYTPYORSQSVFNHFRK
VNEAYOVGLDIDIKRMKNKYGGDGIKOVNFMPSLEYLLSLSEKRFQFGPQIVTL
RPFPEKRLSMNDLENKSHILKFMEOYOKEREAHVSEYLLNLOPCIGADSGKWNPDI
TKLELKSREPDIPLESIRMEFKRYVAATHLKSSKSAKKILOQROANKOELANINN
LMSYLKEVLSSEBONSITYNENINSVNDNGNSKNISDLSYTDKEILEKIVSYIV
DISLYDIENALNAEQLSDNSVDEKTLKRAOSLKLSIMERYAGGRKNDKKNF
FTKDIVGYIMHGISITINTENKNONENPEVHOHNAEENVEHDAEENVEHDAEENVEH
DAEENVEHDAEENVEHDAEENVEENVEENVEENVEENVEENVEENVEENVEENVEE
NVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEE
EVENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEENVEE
801..995
/gene="RESA"
/number=1
996..1198
/gene="RESA"
/number=1
1199..4225
/gene="RESA"
/number=2
BASE COUNT 1933 a 438 c 673 g 1547 t
ORIGIN
Query Match 10.7%; Score 66.4; DB 81; Length 4591;
Best Local Similarity 47.9%; Pred. No.0.074;
Matches 224; Conservative 0; Mismatches 241; Indels 3; Gaps 1;
OY 81 aacatcgataattcttcacgataaagataagatgaagaattgaacaatatgataa 140
DB 3766 ACATGATGCTGAAGAAATGTAGAACATGATGCTGAAGAAATGTAGAAAGAAATGTTGA 3825

OY	141	aaatgtaaaagaacagcgcgagtaaaagataaaaaagcgagcctcaaacctcaattccgaa	200
Db	3626	AGAAAGTGAAGAAAATGTACAAAGAAAATGTAGAGAAAATGTAGAAAATGTTGAAGA	3885
OY	201	agataaatacgaaagtggcagcgtatattgaa---attccagatgctgatatattaagaacc	257
Db	3886	AGTGAAGAAAATGTACAGAAAATGTAGAGAAAATGTAGAAAATGTTGAAGAAA	3945
OY	258	agtatatccagagaccagaacacctgacaataatagagtgtaagctttgcagaaga	317
Db	3946	TGTTGAAGAAAATGTTGMAAAAATGTAGAGAAAATGTAGAGAAAATGTTGAAGAAAT	4005
OY	318	aaatgaatcactgagatgataaaaatttcaatttgagagacacacttattgacgcgc	377
Db	4006	TGATGAAGAAAATGTTGMAAGACTAGAGAAAATGTAGAGAAAATGTTGAAGAAAATGT	4065
OY	378	gaectatcaatlttacaatalttaagcagccaaaagaglaatalgylgtactttaaagt	437
Db	4066	TGAAGAAAATGTAGAACAAAATGTTGAAGAAAGTAGAGAAAATGTTGAAGAAAATGTAGA	4125
OY	438	tggttaatgaaacacgctaagatataaalygaaglataaagagtltlaagcctacaagatlt	497
Db	4126	AGAAAATGTGAAGAGCAATGTTGAGAGAAATGTTGAAGAAATGTTGAAGAAATATGATGA	4185
OY	498	agaggttctgagtgaacaaaaagyltaagataaacatcaatcaat	545
Db	4186	AGAAAATGTTGAAGAACCAATATAGAGAAATGTGATGATTAATAAAAAAAT	4233

RESULT	12								
LOCUS	166494/c								
DEFINITION	166494	7218 bp	DNA						28-DEC-1997
ACCESSION	Sequence 14	from patent US 5670367.							
VERSION	166494								
KEYWORDS	166494.1	GI:2724471							
SOURCE									
ORGANISM	Unknown.								
REFERENCE	Unclassified.								
AUTHORS	1 (bases 1 to 7218)								
TITLE	Donner, F., Scheifflinger, F. and Falkner, F. Gunter.								
JOURNAL	Recombinant fowlpox virus								
FEATURES	Patent: US 5670367-A 14 23-SEP-1997;								
SOURCE	location/Qualifiers								
	1..7218								
	/organism="Unknown"								
BASE COUNT	1944 a	1491 c	1486 g	1929 t					
ORIGIN					368	others			

[illegible][illegible]

RESULT	13
LOCUS	AE001373
DEFINITION	AE001373 12029 bp DNA INV 06-NOV-1998
ACCESSION VERSION	Plasmodium falciparum chromosome 2, section 10 of 73 of the complete sequence. AE001373 AE001365.
KEYWORDS	AE001373.1 GI:3645097
SOURCE	malaria parasite P. falciparum.

REFERENCE
AUTHORS
Gardner, M. J., Tetetelin, H., Carucci, D. J., Cummings, L. M., Aray
Eukaryota; Alveolata; Apicomplexa; Haemosporida; Plasmodium
1 (bases 1 to 1029)

TITLE
Chromosome 2 sequence of the human malaria parasite *Plasmodium*

JOURNAL	Science 282 (5391), 1126-1132 (1998)
MEDLINE	99021743
REMARK	Erratum: [[published erratum appears in Science 1998 Dec 4;282(5395):18271]
REFERENCE	2 (bases 1 to 12029)

TITLE Direct Submission
JOURNAL Submitted (02-NOV-1998) The Institute for Genomic Research, 9712
FEATURES Medical Center Drive, Rockville, MD 20814, USA
source Location/Qualifiers
1. 12029

```

gene      /organism="Plasmodium falciparum"
          /db_xref="taxon:5833"
          /chromosome="2"
          3709..5236
gene      /gene="PF0110w"
          join(3709..4318,5079..5236)
CDS

```

```

/ gene="PF30110W"
/ note="predicted by Glimmer3"
/ codon_start=1
/ product="predicted integral membrane protein"
/ protein_id="AAC71812.1"
/ db_xref="GI:3845098"
/ translation="MNLVLTLEIFHTNKIYIIIIITVILCYLFLCFSEYVKSISIKINR
EKKMYQIVVEREDVIMKODEKITLNEKSEERLNPTEKQIPIYSCSELEKXHNLT
TRVLEIKILLLEQMEVYLIKTDMNMKHKVDEMSWALXEVNFIPIKDLKNEPVV
SYEKETETIQKWFINTSHDYNVPLNVEFERKKHKVKSQCEYVELLTHICSFLFFLLIS
LFSCIFLYLFLPFCIMFVYLLPCLFLINFINKPM"
6675. .10253

```

CDS

```
/gene="PF80115w"
6675..10253
/gene="PF80115w"
/note="predicted by GlimmerM"
/codon_start=1
/product="predicted secreted protein"
/protein_id="AAC71813.1"
/db_xref="GI:3845099"
/translation="MKKRYIIYIGILNFTKNNVSVPEPFLSONKDSFEKKY
TYGNLIQASITNTPKTSQENKDIKEKTKNTIKTNNPESPLNKRFPKLIQLTNRK
ENKDEONKENRKDEONKEDONKONKEDONKONKONKINVSXKLSGNNOQNSIP
KSIOCPENCVKKOSNOPRSPYEPFEANFQPIDLMJETDYSDDLEDOLNYGKRIE
HIDDETQYIGSILGYSDDNNKMKYONTOLIDNNKGKTTMTMEKKNRDKKSKKKR
TKQNYKKKENQNIENIIPSKYKQENIETLIDNNKILKSKNTKEKGIKERTDTTN
IADRIKKEERETKDEKKNITQOLVKNQVOLIKGEEFKDKDEKESKTDDEEDTDEE
```



```

* 10585 11303: contig of 719 bp in length
* 11304 11403: gap of 100 bp
* 11404 12111: contig of 708 bp in length
* 12112 12211: gap of 100 bp
* 12212 12923: contig of 712 bp in length
* 12924 13023: gap of 100 bp
* 13024 13737: contig of 714 bp in length
* 13738 13837: gap of 100 bp
* 13838 14547: contig of 710 bp in length
* 14548 14647: gap of 100 bp
* 14648 15337: contig of 690 bp in length
* 15338 15437: gap of 100 bp
* 15438 16151: contig of 714 bp in length
* 16152 16251: gap of 100 bp
* 16252 16957: contig of 706 bp in length
* 16958 17057: gap of 100 bp
* 17058 17730: contig of 673 bp in length
* 17731 17830: gap of 100 bp
* 17831 18512: contig of 682 bp in length
* 18513 18612: gap of 100 bp
* 18613 19313: contig of 701 bp in length
* 19314 19413: gap of 100 bp
* 19414 20051: contig of 638 bp in length
* 20052 20151: gap of 100 bp
* 20152 20872: contig of 721 bp in length
* 20873 20972: gap of 100 bp
* 20973 21700: contig of 728 bp in length
* 21701 21800: gap of 100 bp
* 21801 22526: contig of 726 bp in length
* 22527 22626: gap of 100 bp
* 22627 23342: contig of 716 bp in length
* 23343 23442: gap of 100 bp
* 23443 24164: contig of 722 bp in length
* 24165 24264: gap of 100 bp
* 24265 24983: contig of 719 bp in length
* 24984 25083: gap of 100 bp
* 25084 25793: contig of 710 bp in length
* 25794 25893: gap of 100 bp
* 25894 26593: contig of 700 bp in length
* 26594 26693: gap of 100 bp
* 26694 27384: contig of 691 bp in length
* 27385 27484: gap of 100 bp
* 27485 28207: contig of 723 bp in length
* 28208 28307: gap of 100 bp
* 28308 29013: contig of 706 bp in length
* 29014 29113: gap of 100 bp
* 29114 29830: contig of 717 bp in length
* 29831 29930: gap of 100 bp
* 29931 30642: contig of 712 bp in length
* 30643 30742: gap of 100 bp
* 30743 31450: contig of 708 bp in length
* 31451 31550: gap of 100 bp
* 31551 32269: contig of 719 bp in length
* 32270 32369: gap of 100 bp
* 32370 33071: contig of 702 bp in length
* 33072 33171: gap of 100 bp
* 33172 33875: contig of 704 bp in length
* 33876 33975: gap of 100 bp
* 33976 34649: contig of 674 bp in length
* 34650 34749: gap of 100 bp
* 34750 35466: contig of 717 bp in length
* 35467 35566: gap of 100 bp
* 35567 36288: contig of 722 bp in length
* 36289 36388: gap of 100 bp
* 36389 37121: contig of 733 bp in length
* 37122 37221: gap of 100 bp
* 37222 37928: contig of 708 bp in length
* 37929 38029: gap of 100 bp
* 38030 38740: contig of 711 bp in length
* 38741 38840: gap of 100 bp
* 38841 39549: contig of 709 bp in length
* 39550 39649: gap of 100 bp
* 39650 40340: contig of 691 bp in length

```

```

* 40341 40440: gap of 100 bp
* 40441 41164: contig of 724 bp in length
* 41165 41264: gap of 100 bp
* 41265 41957: contig of 693 bp in length
* 41958 42057: gap of 100 bp
* 42058 42739: contig of 682 bp in length
* 42740 42839: gap of 100 bp
* 42840 43571: contig of 732 bp in length
* 43572 43671: gap of 100 bp
* 43672 44390: contig of 719 bp in length
* 44391 44490: gap of 100 bp
* 44491 45208: contig of 719 bp in length
* 45210 45309: gap of 100 bp
* 45310 46033: contig of 724 bp in length
* 46034 46133: gap of 100 bp
* 46134 46841: contig of 708 bp in length
* 46842 46941: gap of 100 bp
* 46942 47618: contig of 677 bp in length
* 47619 47718: gap of 100 bp
* 47719 48439: contig of 721 bp in length
* 48440 48539: gap of 100 bp
* 48540 49193: contig of 654 bp in length
* 49194 49293: gap of 100 bp
* 49294 49950: contig of 657 bp in length
* 49951 50050: gap of 100 bp
* 50051 50765: contig of 715 bp in length
* 50766 50865: gap of 100 bp
* 50866 51570: contig of 705 bp in length
* 51571 51670: gap of 100 bp
* 51671 52408: contig of 738 bp in length
* 52409 52508: gap of 100 bp
* 52509 53185: contig of 677 bp in length
* 53186 53285: gap of 100 bp
* 53286 54003: contig of 718 bp in length
* 54004 54103: gap of 100 bp
* 54104 54814: contig of 711 bp in length
* 54815 54914: gap of 100 bp
* 54915 55643: contig of 729 bp in length
* 55644 55743: gap of 100 bp
* 55744 56475: contig of 732 bp in length
* 56476 56575: gap of 100 bp
* 56576 57279: contig of 704 bp in length

Query Match      10.3%  Score 64.2; DB 49; Length 201289;
Best Local Similarity 39.0%; Pred. No. 0.08;
Matches 201; Conservative 0; Mismatches 315; Indels 0; Gaps 0;

Oy 106 aagataagatgaaagatgacaataatgtaaaatgtaagaacagcgagtaaa 165
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159388 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159329

Oy 166 gataaagacgaagcgaacccaattccgaagaataatggaagtgcgagctat 225
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159328 AAAAAAAAAAGGAGANNANNNANNNANNNANNNANNNANNNANNNAN 159269

Oy 226 atgcaattccagatgctgatatataagaaccagatatccagagcgaacactgaa 285
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159268 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159209

Oy 286 caatataatagagtgtaagcttgcaagaagaataatgaactagatgcaaatctt 345
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159208 NAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159149

Oy 346 tcaattgcagacgaaccttcattgacgcctccgaactcaattacgaatcttaagca 405
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159148 AANNNNNNAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159089

Oy 406 gccaaaaaggtagatgltgltacttaagttggtlaatgaacacgtaagtataaatg 465
      ||||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||
Db 159088 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 159029

Oy 466 acaagataagagatgttaagcctacagatgtaagagttcagatgaaacaaagtaaa 525
      ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| ||| |||

```


THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM nucleic - nucleic search, using sw model

Run on: March 14, 2001, 02:37:09 ; Search time 975.78 Seconds
(without alignments)
4459.663 Million cell updates/sec

Title: US-09-292-437-2
Perfect score: 621
Sequence: 1 atgaaaaaatggacaatcg.....tagctacagaagtcacataa 621

Scoring table: IDENTITY_NUC
Gapop 10.0 , Gapext 1.0

Searched: 7991742 segs, 3503743858 residues
Total number of hits satisfying chosen parameters: 15983484

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%
Listing first 45 summaries

Database :

EST:*

1: gb_est1:*
2: gb_est2:*
3: gb_est3:*
4: gb_est4:*
5: gb_est5:*
6: gb_est6:*
7: gb_est7:*
8: gb_est8:*
9: gb_est9:*
10: gb_est10:*
11: gb_est11:*
12: gb_est12:*
13: gb_est13:*
14: gb_est14:*
15: gb_est15:*
16: gb_est16:*
17: gb_est17:*
18: gb_est18:*
19: gb_est19:*
20: gb_est20:*
21: gb_est21:*
22: gb_est22:*
23: gb_est23:*
24: gb_est24:*
25: gb_est25:*
26: gb_est26:*
27: gb_est27:*
28: gb_est28:*
29: gb_est29:*
30: gb_est30:*
31: gb_est31:*
32: gb_est32:*
33: gb_est33:*
34: gb_est34:*
35: gb_est35:*
36: gb_est36:*
37: gb_est37:*
38: gb_est38:*
39: gb_est39:*
40: gb_est40:*
41: em_estba:*
42: em_estfun:*
43: em_esthum1:*

44: em_esthum2:*
45: em_esthum3:*
46: em_esthum4:*
47: em_esthum5:*
48: em_esthum6:*
49: em_esthum7:*
50: em_esthum8:*
51: em_esthum9:*
52: em_esthum10:*
53: em_esthum11:*
54: em_esthum12:*
55: em_esthum13:*
56: em_esthum14:*
57: em_esthum15:*
58: em_esthum16:*
59: em_esthum17:*
60: em_esthum18:*
61: em_esthum19:*
62: em_esthum20:*
63: em_estin1:*
64: em_estin2:*
65: em_estin3:*
66: em_estin4:*
67: em_estlov1:*
68: em_estlov2:*
69: em_estpl1:*
70: em_estpl2:*
71: em_estpl3:*
72: em_estpl4:*
73: em_estpl5:*
74: em_estro1:*
75: em_estro2:*
76: em_estro3:*
77: em_estro4:*
78: em_estro5:*
79: em_estro6:*
80: em_estro7:*
81: em_estro8:*
82: em_estro9:*
83: em_estro10:*
84: em_estro11:*
85: em_estro12:*
86: em_estro13:*
87: gb_esta1:*
88: gb_esta2:*
89: gb_esta3:*
90: gb_esta4:*
91: gb_esta5:*
92: gb_esta6:*
93: gb_esta7:*
94: gb_esta8:*
95: gb_esta9:*
96: gb_esta10:*
97: gb_esta11:*
98: gb_esta12:*
99: gb_esta13:*
100: gb_esta14:*
101: gb_esta15:*
102: gb_esta16:*
103: gb_esta17:*
104: gb_esta18:*
105: gb_esta19:*
106: gb_esta20:*
107: gb_esta21:*
108: gb_esta22:*
109: gb_esta23:*
110: gb_esta24:*
111: gb_esta25:*
112: em_esthum21:*
113: em_esthum22:*
114: em_esthum23:*
115: em_estom1:*
116: em_estom2:*

Result	No.	Score	Match	Query Length	DB	ID	Description
C	1	74.2	11.9	942	190	CNS016GS	AL109318 Drosophila
	2	70.6	11.4	593	190	CNS00860	AL051540 Drosophila
	3	70.6	11.4	894	105	BE215316	BE215316 HV_CEP0000
	4	70.6	11.4	997	190	CNS0134P	AL102403 Drosophila
C	5	69.2	11.1	581	191	CNS010PZ	AL168128 Tetradodon
	6	69.2	11.1	942	190	CNS004H7	AL073768 Drosophila
	7	69	11.1	959	190	CNS00655	AL062806 Drosophila
	8	69	11.1	1223	183	BI2981	BI2981 T24011-Sp6
C	9	68.8	11.1	633	190	CNS0170D	AL108367 Drosophila
	10	68.8	11.1	1101	190	CNS0153V	AL104965 Drosophila
	11	68.8	11.1	1135	191	CNS03350	AL226115 Tetradodon
	12	68.4	11.0	870	156	AQ330286	AQ330286 nbxb00466
C	13	68.4	11.0	1101	190	CNS01807	AL108721 Drosophila
	14	67.4	10.9	952	172	AQ897460	AQ897460 HS_3134_A
	15	67	10.8	1101	190	CNS00KRH	AL077776 Drosophila
	16	66.8	10.8	796	190	CNS0118D	AL099943 Drosophila
C	17	66.6	10.7	1101	190	CNS00HX9	AL073856 Drosophila
	18	66.2	10.7	850	190	CNS009E7	AL053463 Drosophila
	19	66	10.6	576	191	CNS035N7	AL228940 Tetradodon
	20	66	10.6	963	190	CNS0044L	AL054918 Drosophila
C	21	65.6	10.6	776	190	CNS009BD	AL053563 Drosophila
	22	65.6	10.6	780	110	BE642964	BE642964 Cr12_7_ID
	23	65.6	10.6	833	190	CNS007X3	AL050945 Drosophila
	24	65.4	10.5	834	183	BI2387	BI2387 F21P20-Sp6
C	25	65.4	10.5	1198	183	B08337	B08337 T19P9-Sp6.1
	26	65.4	10.5	1885	108	BE420745	BE420745 HMM002_BO
	27	65.2	10.5	770	176	A2196926	A2196926 SP_1033_E
	28	65.2	10.5	850	176	A2186328	A2186328 SP_1006_E
C	29	65	10.5	890	148	AQ026918	AQ026918 CIT-HSP-2
	30	65	10.5	1101	190	CNS015NV	AL101797 Drosophila
	31	64.8	10.4	1059	190	CNS0022B	AL087713 Drosophila
	32	64.4	10.4	614	190	CNS0152H	AL104915 Drosophila
C	33	64.4	10.4	888	191	CNS027EX	AL114578 Tetradodon
	34	64.4	10.4	1101	190	CNS001FB	AL060732 Drosophila
	35	64.4	10.4	1101	190	CNS0039R	AL063932 Drosophila
	36	64.4	10.4	1101	190	CNS017OV	AL108365 Drosophila
C	37	64.2	10.3	730	171	AQ879158	AQ879158 HS_3184_E
	38	64.2	10.3	741	190	CNS007YN	AL051001 Drosophila
	39	64.2	10.3	815	171	AQ853920	AQ853920 nbxb00466
	40	64.2	10.3	865	151	AQ324474	AQ324474 mxh0001887
C	41	64.2	10.3	1101	190	CNS00OLT	AL068307 Drosophila
	42	64	10.3	662	192	CNS0419K	AL291955 Tetradodon
	43	64	10.3	693	169	AQ782441	AQ782441 HS_3174_A
	44	64	10.3	1042	190	CNS018K	AL103838 Drosophila
C	45	63.8	10.3	736	190	CNS009DE	AL053636 Drosophila

ACCESSION	fly), genomic survey s
VERSION	AL109318
KEYWORDS	AL109318.1 GT:56236222
	GSS.

SOURCE	fruit fly.
ORGANISM	Plasmid Drosophila melanogaster
REFERENCE	Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
AUTHORS	Ptergota; Neoptera; Endopterygota; Diptera; Brachycera;
TITLE	Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
JOURNAL	1 (bases 1 to 942)
COMMENT	Genoscope. Direct Submission Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage : BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr - Web : www.genoscope.cns.fr) determination of this BAC-end sequence was carried out as part of a collaboration with the European Drosophila Genome project (EDGP) - http://www.edgp.ebi.ac.uk/. This Drosophila melanogaster BAC library (Dros BAC) was made by Alain Billard at CDPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.
FEATURES	Location/Qualifiers 1..942 /organism="Drosophila melanogaster" /plasmid="pbelobAC11" /db_xref="taxon:71227" /clone_1lb="DrosBAC" /clone="BACN13P09" /note="end : 17"
BASE COUNT	155 a 77 c 70 g 494 t 146 others
ORIGIN	
Query Match	11.9% Score 74.2; DB 190; Length 942;
Best Local Similarity	38.0%; Pred. No. 1.5e-06;
Matches 178; Conservative 52; Mismatches 238; Indels 0; Gaps 0;	
OY	107 aagatgaagaatgaaaggcttgacaatatgtataaaaagttaagaacacggcagtgtaaag 166
Db	575 AACAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAGAAAAAAATTGGAACACMAAC 516
OY	167 ataaaaagcagcaagctlaaacctcaaatccgaaagataaatcgaaagtgcagcgtata 226
Db	515 GAAMAAAAAAAAAACGTAAAMWMMWMAAACAGARARAANAANNNNNANNAANAAAA 456
OY	227 ttgaattccagatgctgatattaagaaccagtatatccaggaccgaacacctgaac 286
Db	455 NAAAAAAAAAAAAAAAAACTRAAMRAATAARAANAAMAAAAMCMRARAFAAAAAAW 396
OY	287 aataatagaagtgtaagcttgcagaagaataagatcactatagatcataaatattc 346
Db	395 AACCGGTARAKMTWTAAATGTAACTCRRAAAATTAIAAAAAAAAAAAAAAAAAAMRTAA 336
OY	347 caattgcagcacacctlcatltgacgcgccgaactlcaatltaacaatctaagcag 406
Db	335 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAMAMMWTTCMCWMAAAAAAAAAAAAA 276
OY	407 ccaaaaagatgtagtctgcttactttaaagttgtaatgaacacgtaagtataaatga 466
Db	275 AAAAAAAAAAAAAAAAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAAAMAA 216
OY	467 caagataagaagtgtaagccttaagaagtgtagaggttcctgatagaacaaaaagtgaa 526
Db	215 WTGAAMAMAMATMGCTTAAMAAATCGTTAMCAAAAMAMAAAMAAAMAAAMAAATWMAAA 156
OY	527 ataacaataacatlaacttacttgatlgatlaactgaacgaagaagcag 574
Db	155 AMGCCAAATTTATGCAATTWAAAAAMAAAMAAAMAAAMAAAMAAAMAAAG 108
RESULT 2	
CNS00880	593 bp DNA GSS 03-JUN-1999
LOCUS	Drosophila melanogaster genome survey sequence TE13 end of BAC #
DEFINITION	BACR16J23 of RFL1-98 library from Drosophila melanogaster [fruit

ACCESSION VERSION KEYWORDS SOURCE ORGANISM	 fly), genomic survey sequence. AL051540 GI:4933381 GSS: fruit fly, Drosophila melanogaster Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophilla.
REFERENCE AUTHORS TITLE JOURNAL	 Genoscope. Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : sequef@genoscope.cns.fr) Web : www.genoscope.cns.fr)
COMMENT	 Determination of this BAC-end sequence was carried out as part of a collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the Drosophila melanogaster genome using these BACS. For further information please see http://www.fruitfly.org The BDGP Drosophila melanogaster BAC library was prepared by Kazutoyo Osoegawa and Aaron Mamoser at Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named ROCI-98 and was constructed by partial EcoRI digestion of Drosophila DNA provided by the BDGP from the isogenic strain y2; cn bw sp, the same strain used for the BGCP's P1 and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACP/C Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.
FEATURES	
source	 1..593 /organism="Drosophila melanogaster" /db_xref="taxon:7227" /clone_lib="RPCI-98" /clone="BACRIC6T23" /note="end : TET3"
BASE COUNT	 448 a 25 c 18 g 10 t 92 others
ORIGIN	
Query Match	 Best Local Similarity 11.4%; Score 70.6; DB 190; Length 593;
Matches 209; Conservative 56; Mismatches 286; Indels 0; Gaps 0;	
Dy 71	tgtctaaaccatcatgatatatttccagcagaataagaatagaagtgaacctgaac 130
Dy 6	TCCDAAAATACAGMGAIAAATAATCAACAATAATAATAAACMGCGAAAAAAAACMA 65
OY 131	aataatgaataaatgttaaaaagacaaggcgtagtaaagataaaaaagcaaacctaacttc 190
Dy 66	AAAAAAAAAMWMMVMAAAAAAAAAAAAAAAAAAATAACCMAMMVMCAMAAMMAAAAAA 125
OY 191	aaattccgaagaataatcgaaatgggcgcgatataatgaataatccagatgcttatatta 250
Dy 126	MAGGACGACAAAAMMAAMMMAACAAAAAACCACAAAAAAMAAAAAACRRRA 185
OY 251	aagaacgatlatatccagaccagcacgaacttgaaatlaaatagaggtgtaagctttg 310
Dy 186	AAAG 245
OY 311	cagaagaataatgaatcagatagtgtaataatattccaatgccagagacgaacttcattg 370
Dy 246	AGAACRMMAAMMAAAAAAAAAAMMRMAAAAAAAAAAAAAAAAAAAAAAAAAAMMAAACMAAM 305
OY 371	accgtlccgaactatcataatcttaaaagcagcaaanaagtgtagtgtgcttac 430
Dy 306	AAMMARMAAMMRRAAAAAAAAAARRRAMAAAAARMAAMGGCAAAGAAGMMA 365
OY 431	ttaaagltggtaagaaacagtgtaagttaaatgacagatataagaagatgtttaagcta 490
Dy 366	ARAAAAMMAAAAAAAAAATMAAGACAATTMAAAAAAAAAAAAAAAAAAAAAAAAACAARAANA 425

```

Oy 491 cagatgtagaggtctcgtatgaacaaaggtaagaataaacaattacaattactt 550
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 426 AMCAAAAAMAAAAMCGAAGGCGCAAAAAMAAAAMAAAAMAAAAMAAAAMAAA 485
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 551 gtgtgtattcaatgaaagcagcgcttgggaaacgctaaacctttagctaacg 610
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 486 AAAAAAAMAAAACGAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAM 545
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 611 aagccaataa 621
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 546 AAAAAAAMAAAAM 556

RESULT 3
LOCUS BE215316 894 bp mRNA EST 25-JUL-2000
DEFINITION HV_CEB0006H11f Hordeum vulgare seedling green leaf EST library
            HV_CEB0006H11f, mRNA sequence.
ACCESSION BE215316
VERSION BE215316.1 GI:8902928
KEYWORDS EST.
SOURCE barley.
ORGANISM Hordeum vulgare
           Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
           Magnoliophyta; Liliopsida; Poales; Poaceae; Hordeum.
REFERENCE 1 (bases 1 to 894)
AUTHORS Wing, R., Close, T.J., Kleinhofs, A., Wise, R., Begum, D., Frisch, D., Yu
           X., Anderson, H., Dale, J., Henry, D., Kernodle, S., Palmer, M., Rambo
           J., Saski, C., Schwartzbeck, V., Stimson, J., Choi, D.W., Main, D. and
           Wood, T.
           Development of a genetically and physically anchored EST resource
           for barley genomics
           Unpublished (2000)
           Contact: Wing RA
           Clemson University Genomics Institute
           Clemson University
           100 Jordan Hall, Clemson, SC 29634, USA
           Tel: 864 656 7288
           Fax: 864 656 4293
           Email: rwing@clemson.edu
           Seq primer: AATTACCTCCTCCTAAGCG
           High quality sequence start: 21.
           High quality sequence stop: 871.
           Location/Qualifiers
             1..894
               /organism="Hordeum vulgare"
               /cultivar="C116151 (M1a6)"
               /db_xref="taxon:4513"
               /clone="HV_CEB0006H11f"
               /clone_11b="Hordeum vulgare seedling green leaf EST
               library HVCDNA0005 (Erysiphe infected & control)"
               /tissue_type="seedling green leaf"
               /lab_host="SOLR"
               /note="Vector: lambdaZAP; Site_1: EcoRI; Site_2: XhoI"

BASE COUNT 724 a 40 c 85 t 4 others
ORIGIN
Query Match 11.4%; Score 70.6; DB 105; Length 894;
Best Local Similarity 46.5%; Pred. No. 9.3e-06;
Matches 229; Conservative 0; Mismatches 264; Indels 0; Gaps 0;

Oy 104 ataagataagaatgaatgataatgataatgataatgataatgataatgataatg 163
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 256 AGAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 315
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 164 aagataaaagcagcaagctaaacctcaatccgaaagataatcgaagctggcagct 223
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 316 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 375
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 224 atattgaattcagatgctgatatataagaaccagatattccaggaccagcaacctg 283
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

Db 376 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 435
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 284 aacaataatagaggtgtaagcttgcagaagaataatgaatcactagatgataaata 343
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 436 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 495
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 344 ttcaattgcagacacacttcatgacgctcgaactatcaatcacaattcaag 403
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 496 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 555
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 404 cagccaaaaggtgtagtctgctttaaagttggtgaatgaacacgtgaataaa 463
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 556 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 615
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 464 tgacaagtataagagtgtaagcctacagatgtagtgagttgatacaaaaagta 523
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 616 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 675
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 524 aagataaacaattacaattacttgatgataatgaagaacagcgcttggg 583
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 676 AATAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 735
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Oy 584 aaaaacgtaaaat 596
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 736 AAAAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAAAMAAA 748
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

RESULT 4
LOCUS CNS0134P/c 997 bp DNA GSS 26-JUL-1999
DEFINITION Drosophila melanogaster genome survey sequence SP6 end of BAC
            BACN09C07 of DrosBAC library from Drosophila melanogaster (fruit
            fly), genomic survey sequence.
ACCESSION AL102403
VERSION AL102403.1 GI:5614014
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
           Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
           1 (bases 1 to 997)
           Genoscope.
           Direct Submission
           Submitted (23-JUL-1999) Genoscope - Centre National de Sequenceage :
           BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
           - Web : www.genoscope.cns.fr)
           Determination of this BAC-end sequence was carried out as part of a
           collaboration with the European Drosophila Genome Project (FDGP) -
           http://www.edgp.ebi.ac.uk -. This Drosophila melanogaster BAC
           library (Dros BAC) was made by Alain Billaud at CEPH (Centre
           d'Etude du Polymorphisme Humain) with funding provided by a MRC
           project grant. The DNA was prepared from embryos by Alain Bucheton
           and Genevieve Payan. It has been constructed in the vector
           pBelobAC11.
           Location/Qualifiers
             1..997
               /organism="Drosophila melanogaster"
               /plasmid="pBelobAC11"
               /db_xref="taxon:7227"
               /clone_11b="DrosBAC"
               /clone="BACN09C07"
               /note="end : SP6"

BASE COUNT 162 a 115 c 79 g 517 t 124 others
ORIGIN
Query Match 11.4%; Score 70.6; DB 190; Length 997;
Best Local Similarity 39.1%; Pred. No. 9.3e-06;
Matches 185; Conservative 42; Mismatches 243; Indels 3; Gaps 1;

Oy 76 aaaccatataatgataattcttcagataaagataagataagataagataat 135
      : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

Db 549 AAAAAAAAAAAAAAAAAAATRAAAMRTAAAAAAAAAAAAAAAAAATRAAAM 490
Oy 136 gataaataatgtaaaagacagcgagtaaaagtaaaagcagcaagcctaaccataat 195
Db 489 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 430
Oy 196 ccgaagaataatcgaaagtcgagctatattgaaatccagatgctgatatgaaag 255
Db 439 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 373
Oy 256 ccagatattccagagccagacacactgacataatgaaagtgtaagctttgcaaga 315
Db 372 TAAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTAATTA 313
Oy 316 gaaataatgataatgataatgataatgataatgataatgataatgataatgata 375
Db 312 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 253
Oy 376 ccgaactataatgataatgataatgataatgataatgataatgataatgataat 435
Db 252 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 193
Oy 436 gttgtaataagacacgtaagataatgataatgataatgataatgataatgataat 495
Db 132 AATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAAATAA 133
Oy 486 gtagagatctagatgataatgataatgataatgataatgataatgataatgata 548
Db 132 TAAAGCATGATGATGATGATGATGATGATGATGATGATGATGATGATGATGAT 80

RESULT 5

CNS01UPZ 581 bp DNA GSS 12-MAY-2000
LOCUS Tetradon nigroviridis genome survey sequence PUC-ori end of clone
DEFINITION 197M17 of library G from Tetradon nigroviridis, genomic survey
sequence.

ACCESSION AL168128.1 GI:7806185
VERSION GSS: genome survey sequence.
KEYWORDS Tetradon nigroviridis.
SOURCE Tetradon nigroviridis.
ORGANISM Tetradon nigroviridis.
Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;
Eurypterygii; Ctenosquamata; Acanthomorpha; Eucanthomorpha;
Holacanthopterygii; Acanthopterygii; Percomorpha;
Tetraodontiformes; Tetraodontidae; Tetraodon.

REFERENCE 1 (bases 1 to 581)
AUTHORS Roest-Crollius,H., Jallion,O., Dasilva,C., Fizes,C., Fisher,C.,
Bonneau,L., Billault,A., Quetier,F., Saurin,W., Bernot,A. and
Weissenbach,J.

TITLE Characterization and repeat analysis of the compact genome of the
freshwater pufferfish Tetradon nigroviridis
JOURNAL Unpublished
AUTHORS 2 (bases 1 to 581)
Roest-Crollius,H., Jallion,O., Dasilva,C., Bonneau,L., Fisher,C.,
Bernot,A., Fizes,C., Mincker,P., Brothier,P., Quetier,F.,
Saurin,W. and Weissenbach,J.

TITLE Human gene number estimate provided by genome wide analysis using
Tetradon nigroviridis DNA sequence
JOURNAL Unpublished
AUTHORS 3 (bases 1 to 581)
Genoscope.

REFERENCE Direct Submission
JOURNAL Submitted (12-APR-2000) to the EMBL/Genbank/DBJ databases
COMMENT This sequence is a single read and was generated as part of a large
genome. For more information, please take a look at
http://www.genoscope.cns.fr/Tetradon.

FEATURES
Source
1..581
/organism="Tetradon nigroviridis"
/db_xref="taxon:99883"

/clone="197M17"
/clone_lib="G"
/note="Genoscope sequence ID : C0AG197AG09SP1-end :
PUC-ori"
BASE COUNT 417 a 7 c 58 g 31 t 68 others
ORIGIN

Query Match 11.1%; Score 69.2; DB 191; Length 581;
Best Local Similarity 41.7%; Pred. No. 1.8e-05;
Matches 206; Conservative 26; Mismatches 260; Indels 2; Gaps 1;

Oy 104 ataaagataaagataaagattgaaacataatgataaataatgaaagacagcgagta 163
Db 80 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 139
Oy 164 aagataaagacgacagcctaaccctcaatctcgaagaataatcgaaagtcgagct 223
Db 140 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 199
Oy 224 atattgaattccagatgctgataatgataatgataatgataatgataatgataat 283
Db 200 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 259
Oy 284 aacataataatagagtgtaagcttgcagaagaataatgataatgataatgataat 343
Db 260 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 319
Oy 344 ttcaattgcagacacacttcaatgacccgtccgaactcaatgataatgataatg 403
Db 320 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 379
Oy 404 cagcacaagaagtgatgagtggt--acttaaaagtgtgtaagacagcagtaagtaa 461
Db 380 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 439
Oy 462 aatgacaagataagatgtaagcttgaagctacagatgtaagcttgaatgataaag 521
Db 440 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 499
Oy 522 taagaataaacaattacataatgataatgataatgataatgataatgataatg 581
Db 500 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 559
Oy 582 ggaagaacgtaaaa 595
Db 560 AAAAAAAAAAAGMAATTAATAATAAAAAAAAAAAAAAAAAAAGATTAATAATAA 573

RESULT 6
LOCUS CNS00HP7 942 bp DNA GSS 03-JUN-1999
DEFINITION Drosophila melanogaster genome survey sequence T7 end of BAC:
BACR35J09 of RPCI-98 library from Drosophila melanogaster (fruit
fly), genomic survey sequence.

ACCESSION AL073768.1 GI:4953548
VERSION AL073768
KEYWORDS GSS.
SOURCE fruit fly.
ORGANISM Drosophila melanogaster
Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
Pterygota; Neoptera; Endopterygota; Diptera; Brachycera;
Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.

REFERENCE 1 (bases 1 to 942)
Genoscope.

REFERENCE Direct Submission
JOURNAL Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage :
BP 191 91006 EVRY cedex - FRANCE (E-mail : segref@genoscope.cns.fr
- Web : www.genoscope.cns.fr)
COMMENT determination of this BAC-end sequence was carried out as part of a
collaboration with the Berkeley Drosophila Genome Project (BDGP).
The BDGP is constructing a physical map of the Drosophila
melanogaster genome using these BACs. For further information


```

RESULT      8
Bi2981/c    1223 bp    DNA    GSS    14-MAY-1997
LOCUS      T24D11-sp6 Arabidopsis thaliana genomic clone T24D11, DNA
DEFINITION sequence.
ACCESSION  BI2981.1 GI:2094103
VERSION    BI2981.1
KEYWORDS   GSS.
SOURCE     Arabidopsis thaliana
           thale cress.
           Arabidopsis thaliana
           Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
           Magnoliophyta; eudicotyledons; core eudicot; Rosidae; eurosids II;
           Brassicales; Brassicaceae; Arabidopsids.
           1 (bases 1 to 1223)
REFERENCE  Feng,J., Dewar,K., Buehler,E., Kim,C., Li,Y., Shinn,P., Sun,H. and
AUTHORS    Ecker,J.
TITLE      BAC End Sequences at ATGC
JOURNAL     Unpublished (1997)
COMMENT    Contact: Ecker J.
           Arabidopsis Thaliana Genome Center
           University of Pennsylvania
           Dept. of Biology, University of Pennsylvania, Philadelphia, PA
           19104
           Tel: 215-898-9384
           Fax: 215-898-8780
           Email: jecker@genome.bio.upenn.edu
           Seq primer: Sp6
           Class: BAC ends
           High quality sequence start: 199
           High quality sequence stop: 271.
           Location/Qualifiers
             source          1..1223
                           /organism="Arabidopsis thaliana"
                           /strain="Columbia"
                           /db_xref="taxon:3702"
                           /clone="T24D11"
                           /clone_11b="TAMU"
                           /sex="hermaphrodite"
                           /note="Vector: BelobAC11; Site_1: HindIII; Site_2: HindIII
                           ; Produced by Rod Wing"
BASE COUNT  70 a      50 c      53 g      678 t      372 others
ORIGIN
Query Match      11.1%; Score 69; DB 183; Length 1223;
Best Local Similarity 36.9%; Pred. No. 2.1e-05;
Matches 201; Conservative 0; Mismatches 344; Indels 0; Gaps 0;

QY 76 aaacacatcgtataatcttcacgataaagataaagatgaaagattgacaatat 135
    || || || || || || || || || || || || || || || || || || ||
DB 1145 AANNNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1086
QY 136 gataaaatgtaaaagacagcggttaaagataaaagcagcagcttaacctcaatt 195
    || || || || || || || || || || || || || || || || || || ||
DB 1085 NAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 1026
QY 196 ccgaagaataaactcgaagtgccagcttatattgaaatccagatgctgataaaga 255
    || || || || || || || || || || || || || || || || || || ||
DB 1025 ANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 966
QY 256 ccagtatatccgagcagcagcacacgtgaacaattaaatagaagtgtaagcttgcaga 315
    || || || || || || || || || || || || || || || || || || ||
DB 965 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNN 906
QY 316 gaaatagaatcactagatgataaaatattcaattgctgaggaacaccttcattgcagt 375
    || || || || || || || || || || || || || || || || || || ||
DB 905 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 846
QY 376 ccgaactatcaatttcaactcttaagcagccaaagagtgatgtgtgtaacttaaa 435
    || || || || || || || || || || || || || || || || || || ||
DB 845 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 786
QY 436 gtgtgtaatgaacacgtaagataaaatgacaagataagagatgtaagcctacagat 495

```

```

DB 785 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 726
QY 496 gtaagagcttctagatgacacaaggttaaagataaacaattacattacttgctgt 555
    || || || || || || || || || || || || || || || || || || ||
DB 725 AAAAAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNAANNA 666
QY 556 gattacaatgaaagacagcgcttcttggaataaactcttgtagctacagaagtc 615
    || || || || || || || || || || || || || || || || || || ||
DB 665 NAAAAANNNANANNAANNAANNAANNAANNAANNAANNAANNAANNAANNN 606
QY 616 aaata 620
    || ||
DB 605 NAANA 601
    .

RESULT      9
CNS017QD    639 bp    DNA    GSS    26-JUL-1999
LOCUS      Drosophila melanogaster genome survey sequence SP6 end of BAC
DEFINITION BACN37M13 of DrosBAC library from Drosophila melanogaster (fruit
           fly), genomic survey sequence.
ACCESSION  AL108367.1 GI:5628671
VERSION    AL108367
KEYWORDS   GSS.
SOURCE     Plasmid Drosophila melanogaster
           fruit fly.
ORGANISM   Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
           Pterygota; Neoptera; Endopterygota; Diptera; Brachyceta;
           Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
           1 (bases 1 to 639)
REFERENCE  Direct Submission
AUTHORS     Submitted (23-JUL-1999) Genoscope - Centre National de Sequencage :
TITLE      BP 191 91006 EVRI cedex - FRANCE (E-mail : seqref@genoscope.cns.fr
JOURNAL     - Web : www.genoscope.cns.fr)
           Determination of this BAC-end sequence was carried out as part of a
           collaboration with the European Drosophila genome Project (EDGP) -
           http://www.edgp.ebi.ac.uk - This Drosophila melanogaster BAC
           library (Dros BAC) was made by Alain Billard at CERH (Centre
           d'Etude du Polymorphisme Humain) with funding provided by a MRC
           Project grant. The DNA was prepared from embryos by Alain Bucheton
           and Genevieve Payan. It has been constructed in the vector
           pBelobAC11.
FEATURES
  source          Location/Qualifiers
    1..639
    /organism="Drosophila melanogaster"
    /plasmid="pBelobAC11"
    /db_xref="taxon:7227"
    /clone_11b="DrosBAC"
    /clone="BACN37M13"
    /note="end : SP6"
BASE COUNT  389 a      53 c      46 g      52 t      99 others
ORIGIN
Query Match      11.1%; Score 68.8; DB 190; Length 639;
Best Local Similarity 38.3%; Pred. No. 2.2e-05;
Matches 185; Conservative 51; Mismatches 247; Indels 0; Gaps 0;

QY 91 aattatcttcacgataaagaataaagataaagattgaacaatatgataaaatgtaaa 150
    || || || || || || || || || || || || || || || || || || ||
DB 140 AAAAAATTGAARARACAAAMAYAAAHWRRAAMWTWAAAAAGAAAAAATVAAAAARAA 199
QY 151 gaacagcgagtaagaataaagaacagcagcgaactaaacctcaattccgaagataatcg 210
    || || || || || || || || || || || || || || || || || || ||
DB 200 GTTAGGAGGATTAAMWTAAAMRAAAMAAAMAAAMATATCAACMWAARFAAATTA 259
QY 211 aagatgagcgatatttgaattccgagtgctgatatlaaagaacagatatccagga 270
    || || || || || || || || || || || || || || || || || || ||
DB 260 ARAAMWATGAGACTAAWAAAMATWTAAWRTTTTAAGAAAAAAMWGTGAWMARAR 319
QY 271 ccagcaacacctgaaacaattaaatagaagtgtaagcttgcagaagaataatgaatcacta 330

```

Db 320 ARAAAMGAMWMAWMAAATGAARARARMAAGAGMAAGTCAAAAACACAAAAMAAAAAAGAAR 379

QY 331 gatgatcaaaatatttcaattgcagagcacacttccattgacgcgtccgaactaactt 390

Db 380 AAGAAWMAAWMARARAWMTWGAAGTAAWMAAWRAWGTATTTAAAAAAMAAAAA 439

QY 391 acaactttaaagcagcacaanaagtagtctgtaacttaagtgtgtaaagaaca 450

Db 440 AARAAAMWMAAAACAAAGAAAGAAATGAAAAAAMAAAAAATRAAARAAAAANNA 499

QY 451 cgtatgataaaatcgacaaglataaagatcgttaagccctacagatgtagagttcagat 510

Db 500 CAATGAAATATGCAAGAGGATAAAAAAMAAAAAAGAAAGAAAAAAMAAAAA 559

QY 511 gaacaaaggtlaaagataacaacttaacatlaacttctgtgacatlaacaaatgaag 570

Db 560 AAAAAAAGCAAAAAAAMAAAAAAMAAAAAAMAAAAAAMAAAAATTAANNAAAAAA 619

QY 571 aca 573

Db 620 AAA 622

RESULT 10	LOCUS	DEFINITION
CNS0153V	1101 bp	DNA
CNS0153V		GSS
		26-JUL-1999
		Drosophila melanogaster genome survey sequence T7 end of BAC
		BACN12N03 of DrosBAC library from Drosophila melanogaster (fruit
		fly), genomic survey sequence.

ACCESSION	AL104965
VERSION	AL104965.1
KEYWORDS	GI:5616979 GSS.

ORGANISM	Plasmodium	Drosophila	melanogaster
Eukaryota:	Metazoa:	Arthropoda:	Tracheata: Hexapoda: Insecta:
Pterygota:	Neoptera:	Endopterygota:	Diptera: Brachycera:
Muscomorpha:	Ephydroidea:	Drosophilidae:	Drosophila:
1 (bases 1 to 1101)			
Genoscope:			
Direct Submission			
Submitted (23-JUL-1999)	Genoscope - Centre National de Séquençage		
JOURNAL			

COMMENT

collaboration with the European Drosophila Genome Project (EDGP) - <http://www.edgp.ebi.ac.uk> - This *Drosophila melanogaster* BAC library (Dros BAC) was made by Alain Billaud at CPH (Centre d'Etude du Polymorphisme Humain) with funding provided by a MRC project grant. The DNA was prepared from embryos by Alain Bucheton and Genevieve Payan. It has been constructed in the vector pBelobAC11.

FEATURES	Location/Qualifiers
source	1. .1101

```

BASE COUNT      722 a      61 c      54 g      90 t      174 others
ORIGIN

```

Query Match	11.1%	Score 68.8	DB 190	Length 1101
Best Local Similarity	44.0%	Pred. No. 2.3e-05		
Matches 226	Conservative 10	Mismatches 278	Indels 0	Gaps 0

Qy 81 acacatcgataattctcttcgcgataaagataaagatggaagaattgaacatatgttaa 1400

Db 4 AAANAAANAAANAAA 63

141 aaatgtaaaagacagcgagtaagataaaaagcagcaagctaaccctcaatttcgaa 200

Db

64 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 123

QY 201 agataaatcgaaagtgcagcgtatatltgaattccagatgctgatattaagaaccagt 260

261 ataccagaaacacacactaaacattaaatagaadtatgaactttccaaaaaaa 320

[illegible]

QY 321 tgaatcaactagatgatcaaaaataattcaattgcgcgagacaccttcatgtgaccgtccgaa 380
: || | | . | |||| | : | | | | : | : ||

381 ctatcaattacaacatcttaagcaaccaaaaaagtaqatqatqactttaagttg 440

Db 304 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAATATAAAAAAAAAAAAAAA 363

441 taatgaacacgctaagtatataaatgacacgatatagagatgttaagccctacagatgtagg 500

501 agttctagatgaacaaaaggtaaagataaaccaattaacatlaattacttqtgatgattta 560

Db 424 ANAAAAAAAAAAAAAAAAAAAAAATAAAAAAAAAAAAAAAAAAAATTAATAAAAAAAAAA 483

0y caatgaaaagacagcgcttcgggaaaaaacgtlaaa 594
 || |||| | |||| |
0b || |||| | |||| | E17
 || |||| | |||| | |||| |

DEFINITION	LOCUS	RESULT 11
Tetradodon nigroviridis genome survey sequence pUC-ori end of clone	CNS033GQ	1135 bp DNA GSS 15-MAY-2000

208bp24 of library G from *Tetradodon nigroviridis*, genomic survey sequence.

VERSION	AL226115.1	GI:7885026
KEYWORDS	GSS; genome survey sequence.	

ORGANISM Tetraodon nigroviridis
Eukaryota: Metazoa: Chordata: Craniata: Vertebrata: Euteleostomi:

Enurpterygii; Ctenosquamata; Acanthomorphia; Eucanthomorphia; Neoteleostei; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei; Enurpterygii; Ctenosquamata; Acanthomorphia; Eucanthomorphia; Neoteleostei; Actinopterygii; Neopterygii; Teleostei; Euteleostei; Neoteleostei;

REFERENCE
1 (bases 1 to 1135)
Tetraodontiformes; Tetraodontoidae; Tetraodontidae; Tetraodon.

Bonneau, L., Billault, A., Quetier, F., Saurin, W., Bernot, A. and Weissenbach, J.

JOURNAL of Fish Diseases 2011, 34, 1155-1164
doi:10.1017/S0950268811000605

REFERENCE
2 (bases 1 to 1135)
Roest-Crolius, H., Jalllon, O., Dasilva, C., Bouneau, L., Fisher, C.,

TITLE Saurin, W. and Weissenbach, J.
Human gene number estimate provided by genome wide analysis using

JOURNAL
REFERENCE
3 (bases 1 to 1135)
unpublished
leucococin nigroviridius dna sequence

genoscope.	
AUTHORS	
TITLE	Direct Submission
REMARKS	

COMMENT
This sequence is a single read and was generated as part of a large scale clone-end sequencing project of the Tetraodon nigroviridis

location/qualifiers

```
source      1..1135
            /organism="Tetradon nigroviridis"
```

BASE COUNT	863 a	65 c	43 g	124 t	40 others
ORIGIN	/db_xref="taxon:99883" /clone="208P24" /clone_1bp="c" /note="Genoscope sequence ID : C0AC20BH12SP1-end PUC-ori"				

Query Match	11.1%	Score 68.8;	DB 191;	Length 1135;
Best Local Similarity	45.48;	Pred. No. 2.3e-05;		
Matches 235, Conservative	3;	Mismatches 280;	Indels 0;	Gaps 0

OY	104	ataaagataaagctgaagaaagcttgacaacttgcgtaaagatgtaaagacagcgcgta	153
Db	219	AA	278
OY	164	aagataaagaagcagcaagctaaacctcaattccgaagaataatcgaaagtgcagct	223
Db	279	AA	338
OY	224	atatlgaaattccagatgcgtgatattaaagaccgataatccagagccagcaaccttg	283
Db	339	AA	398
OY	284	aacaattaaatagagctgtaagcttgcagagaanaatlgaaatcctgatatcaataa	343
Db	399	AA	458
OY	344	tttcagtcgagcgacaccttcctgacgcgtccgaactatcaattcaaaccttaag	403
Db	459	AAAAAAAAATTAATAA	518
OY	404	cagccaaaagaagctagtaatgcgttactttaaagcttgctaagaaacgcgtaaatlaaa	463
Db	519	AAAAAAAAAAAAAAAAAAAAAAAAATTTAAAAAATAAAAAAAAAAAATTTAAAAATAAAAA	578
OY	464	tgcacagatataagagatgttaagcctcagatgtagagatcttagatggaaacaaagta	523
Db	579	AAAAAAAAATTAATAA	638
OY	524	aagataaacaattcaactaaatctactgtgtgaatgaatgaagaacagcgcttgcg	583
Db	639	ATTAAAAAATAATTTAAAAAATTTAAAAAATAATTTAAAAAATAATTTAAAAA	698
OY	584	aaaaacgtaaaaactcttgtagctcagaagatcaataa	621
Db	699	AAAAAAAAAAAAAAAAATTTTAAAAAATAAAAAAAAAAAAAAAAAAAAAA	736

RESULT 12
AO330286/c
LOCUS
DEFINITION nbx00406318r CUGI Rice BAC Library Oryza sativa genomic clone
ACCESSION AO330286
VERSION AO330286.1 GI:4122136
KEYWORDS GSS.
SOURCE Oryza sativa.
ORGANISM Oryza sativa
REFERENCE Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta
AUTHORS Magnoliophyta; Liliopsida; Poales; Poaceae; Oryza.
TITLE 1 (bases 1 to 870)
JOURNAL Wing, R.A. and Dean, R.A.
COMMENT A BAC End Sequencing Framework to Sequence the Rice genome
Unpublished (1998)
Contact: Wing RA
Clemson University Genomics Institute
100 Jordan Hall, Clemson, SC 29634, USA
Tel: 864 656 7286
Fax: 864 656 4293
Email: twing@clemson.edu

```
Seq primer: GGAACACAGCTATGACCATC
Class: BAC ends
High quality sequence start: 13
High quality sequence stop: 104
location/Qualifiers
    1..870
```

/lab host="E. coli DH10B"
 /note="Vector: pGLOBAC11; Site 1: HindIII; Site 2:
 HindIII. Rice is one of two most popular grains in the
 world. Half of the world population especially those
 inhabiting highly populated areas of the humid tropics
 and subtropics, rely on rice as their primary source of
 carbohydrate. Monocotyledonous rice is a diploid plant
 (2n=24) with a haploid genome equivalent of 431 Mbp
 (Arumuganathan and Earle, 1991). The relatively small
 genome of rice, three times larger than that of
 Arabidopsis, makes it suitable for genomic studies. In
 order to facilitate positional cloning, physical mapping
 and genome sequencing of rice, we have constructed a BAC
 library from *Oryza sativa*, Nipompare variety. The
 library contains 36,864 clones with an average insert size
 of 128.5 kb providing 10.9 haploid genome equivalents. The
 deep coverage allows the isolation a particular sequence
 with a probability of 99.9 %. Two high density filters,
 each containing 18,432 clones (doubly spotted), represent
 the whole library for colony screening."

Query Match	1.08;	Score 68.4;	DB 156;	Length 870;
Best Local Similarly	42.38;	Pred. No. 2.8e-05;		
Matches 219; Conservative	0;	Mismatches 259;	Indels 0;	Gaps 0;

[illegible]

```

QY      584  aaaaacgcaaatcttctgtagccacagaagtcacaataa  621
          |||  |||  |  |||  |||  |||
Db      283  NAAAAANNNAAAAANNNANNNAAAAANAAAAANAAAAA  246

```

RESULT 13

LOCUS	DEFINITION	GENOME	DATE
CNS01807	1101 bp	DNA	26-JUL-1999
	Drosophila melanogaster genome survey sequence Sp6 end of BAC		
	PAON37F07 of DrosBAC library from Drosophila melanogaster (fruit		
	fly), genomic survey sequence.		
ACCESSION	U008731		

BASE COUNT	179 a	70 c	85 g	448 t	319 others
ORIGIN					

Query Match	11.0%	Score 68.4	DB 190	Length 1101
Best Local Similarity	40.3%	Pred. No. 2.8e-05		
Matches 198; Conservative	25	Mismatches 263	Indels 0	Gaps 0

```
Oy 77aacacatatcgaataattatcttcacgataagaataagaaagtgcacaataty136
      |||::|||:::|||::|||
Db 580AAAAAAAAAAATAAATAAAAAAATAAAAAAATAAAAAATTA 522
```

Qy 137 ataaaaatgtcaaaagacacagcgagtaagaatataaaagcagcaagctaaacctcaaatc 196
 | ||| ||| | | ||| : | | | |||
 Db 520 AAAAAACAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAAATAAAAAA 461

```
QY 197 cgaagataaactcgaaagtgcgagcctatatltgaatccagatgctgatattaagaac 256
      ||| | ||| ||| ||| : ||| : | : | : ||| ||
Db 460 AAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAAA 401
```

Qy 257 cagtatatccaggaccagcacaaccttaacaaltaatatgagtggtgaagcttgcagaag 316
- : : | - - - | : : | : : | : : | : :
Db 400 AAATAWAAATTAATAAATAAAAAAAAAAWPAATTAATAAATAAATAAATAA 341

QY 317 aaaaatgaatcactagatgatcaaaaatatttcaattgcagagacacactttcaattgacgcgc 376
 ||| ||| | | | | : : |||| : ||
 Db 340 AAAAATAATATAAAAAAATATATAWAWAAAAAATTAATATATCAANNACACANTCTTCMNA 281

QY 437 ttgttaatcgaacacgtaagcataaaatlgaacaagtataagaagatgtttaagccctacagatg 496
 : | : | : | : | : | : | : | : | : |
Db 220 AAATTAWAAAAAANAANAAATAAAAATNNNNNNNTNATTTATCTTAAANNNAANAANTATN 161

Oy	557	attacaatgaa	567
Db	100	ATTNNAAAAA	90

BASE COUNT ORIGIN	a	c	g	t	others
720	20	110	32	70	

Query Match	10.9%	Score 67.4	DB 172	Length 952
Best Local Similarity	47.2%	Pred. No. 4.6e-05		
Matches 233, Conservative	0	Mismatches 259	Indels 2	Gaps 1

DQ 104 ataaagatlaaagaatgaaagaattgaacatatgatataaatgtttaaagacaggcgagtta 163
| | | | | | | | | | | | | | | |
Db 349 AGAAANAAAAAAAAACGCAAAAAAAGCGAIAAAAAAAAAAAAAAAAAAGAGA 408

224 atattgaaattccagatgctgatatlaagaaccgtatatccagaccgaacacctg 283

[illegible]

RESULT	15
CNS00KHB/c	
LOCUS	CNS00KHB 1101 bp DNA
DEFINITION	Drosophila melanogaster genome survey sequence T7 end of BAC: BACR17J10 of RPc1-98 library from Drosophila melanogaster (fruit fly), genomic survey sequence.

ACCESSION	VERSION	KEYWORDS	SOURCE	ORGANISM	REFERENCE
AL071116	1	GI:4957152	GSS.	Fruit fly. <i>Drosophila melanogaster</i>	
				Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.	
				1 (bases 1 to 1101)	

<p> AUTHORS TITLE JOURNAL COMMENT </p>	<p> Genoscope, Direct Submission Submitted (02-JUN-1999) Genoscope - Centre National de Sequencage BP 191 91006 EVRY cedex - FRANCE (E-mail : seqref@genoscope.cns.fr) Web : www.genoscope.cns.fr Determination of this BAC-end sequence was carried out as part of </p>
---	---

collaboration with the Berkeley Drosophila Genome Project (BDGP). The BDGP is constructing a physical map of the *Drosophila melanogaster* genome using these BACs. For further information please see <http://www.fruitfly.org>. The BDGP *Drosophila melanogaster* BAC library was prepared by Kazutoyo Oseegawa and Aaron Mammossel in Pieter de Jong's laboratory in the Department of Cancer Genetics at the Roswell Park Cancer Institute in Buffalo, NY. The library is named RPECI-98 and was constructed by partial *EcoRI* digestion of *Drosophila* DNA provided by the BDGP from the isogenic strain Y2: cn bw sp, the same strain used for the BDGP-PI and EST libraries. A more detailed description of the library and how to order individual BAC clones, the entire library, or filters for hybridization from the BACPAC Resource Center can be found at http://bacpac.med.buffalo.edu/drosophila_bac.htm.

FEATURES	source	Location/Qualifiers
		1. 1101
		/organism="Drosophila melanogaster"
		/db_xref="taxon:7227"
		/clone_lib="RP11-98"
		/clone="BACR17J10"
		/note="end : 17"
BASE COUNT	170 a	142 c 122 g 590 t
ORIGIN		77 others

Query Match	10.8%	Score 67	DB 190	Length 1101
Best Local Similarity	42.4%	Pred. No.	5.7e-05	
Matches 211	Conservative 23	Mismatches 264	Indels 0	Gaps 0

[illegible]

Search completed: March 14, 2001, 14:04:33
Job time: 41244 sec

Job time: 41244 sec

THIS PAGE BLANK (USPTO)


```

0Y 30 DNYLHDKRDEKLEIDYDANNAEQASKDKKQAKPOIPKDKSVAGYIELPADIKEPYR 89
Dd 5 DDYFSDFDEDELDLKYKAEKANKSVKETTITRRA---VPYOK-----DLHDNYLP 49
0Y 90 GPAPF-EOLNRCVSAEENESLDDONISLAGH-TPIDPNPYOTNL-KAAKGSVNYEKV 146
Dd 50 GOKTYEEELQRDVSGPPIHHLEIDYDALSFYIPIVNEYRVDYOTYIVKHSLFONTLCIPT 109
0Y 147 G-----NETRRYKMTSIRDVPTDVGVLDECK-----GKROKOLLTYICD-DY 187
Dd 110 GMGKFTIASYWLNLNFRRTTKKAKIIFTPTPRIVAOQKACLGTIGISPDQTAILLDSKR 169
0Y 188 NEKTGVWEKKRIEYAT 203
Dd 170 KNREIMANKRVEFAT 185

RESULT 4
NFM_RAT ID NFM_RAT STANDARD; PRT; 845 AA.
AC P12839;
DT 01-OCT-1989 (Rel. 12, Created)
DT 01-MAY-1992 (Rel. 22, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NF-M).
GN NEFM OR NEFM.
OS Rattus norvegicus (Rat).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Rodentia; Sciurognathi; Muridae; Murinae; Rattus.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=87282618; PubMed=2441012;
RA Napoliitano E.W., Chin S.S.M., Colman D.R., Liem R.K.H.;
RT "Complete amino acid sequence and in vitro expression of rat NF-M,
RL the middle molecular weight neurofilament protein.";
RL J. Neurosci. 7:2590-2599(1987).
RN [2]
RP PHOSPHORYLATION SITES, AND REVISION TO 500.
RX MEDLINE=92165797; PubMed=1537832;
RA Xu Z.-S., Liu W.-S., Willard M.B.;
RT "Identification of six phosphorylation sites in the COOH-terminal
RL tail region of the rat neurofilament protein M.";
RL J. Biol. Chem. 267:4467-4471(1992).
RN [3]
RP CARBOHYDRATE-LINKAGE SITES.
RX MEDLINE=93346421; PubMed=8344946;
RA Dong D.L.-Y., Xu Z.-S., Chevrier M.R., Cotter R.J., Cleveland D.W.,
RA Hart G.W.;
RT "Glycosylation of mammalian neurofilaments. Localization of multiple
RL O-linked N-acetylglucosamine moieties on neurofilament polypeptides
RL L and M.";
RL J. Biol. Chem. 268:16679-16687(1993).
CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
CC -I- PWM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETIDE K-S-P. NEFM IS
CC PHOSPHORYLATED ON A NUMBER OF THE SRINES IN THIS MOTIF. IT IS
CC THOUGHT THAT PHOSPHORYLATION OF NEF RESULTS IN THE FORMATION OF
CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
CC OF AXONAL CALIBER.
CC -I- PWM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NF-M AND NF-H), THE
CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
CC -----
CC CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL Outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
```

CC	-----	
DR	EMBL; M18628; AAA1696.1; -.	
DR	INTERPRO; IPR001664; -.	
DR	PFAM; PF00038; filament; 1.	
DR	PROSITE; PS00226; IF; 1.	
KW	Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;	
KW	Phosphorylation; Glycoprotein.	
FT	INIT_METT	0
FT	DOMAIN	1
FT	DOMAIN	103
FT	DOMAIN	103
FT	DOMAIN	411
FT	DOMAIN	103
FT	DOMAIN	135
FT	DOMAIN	148
FT	DOMAIN	247
FT	DOMAIN	264
FT	DOMAIN	285
FT	DOMAIN	286
FT	DOMAIN	290
FT	CARBOHYD	47
FT	CARBOHYD	430
FT	MOD_RES	502
FT	MOD_RES	506
FT	MOD_RES	536
FT	MOD_RES	603
FT	MOD_RES	608
FT	MOD_RES	666
FT	MOD_RES	666
FT	CONFLICT	500
SO	SEQUENCE	845 AA; 95660 MW; 316C41655B1197D CRC64;

```

Query Match          8.5%: Score 91.5; DB 1; Length 845;
Best Local Similarity 23.7%: Pred. No. 10;
Matches 44: Conservative 29; Mismatches 54; Indels 59; Gaps 9;

OY 26 KPHINYLHDKRDKDKIEQDYDNKVKESKDKQOAKPOIPKDSKVAGYIEIPD----- 80
    || ::::|::|: | |::|::|: | |::|::|: | |::|::|: | |::|::|: | |
DB 616 KPEAKAGNDEQKAEKRVLEK-KKEVAKESKPEKEKVEKKEKPEPD-----VPOKKKAE 665
OY 81 ADIKEPVYPGPAT-----PEOLNRGVSAFAE-----NESLDD----- 112
    |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 666 SPVKKAEVEMITITIKSVKSLSEKDTKEEKPOQOEKVKKEKAEDEEGSEBEVDGSKSPQESK 725
OY 113 -QNISAGHTFTIDRNRNYQFTNLKAAKKGSMYVFKXGNENRKKMHSIRYKRTDGVLD 171
    ::::|::|: |::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|::|
DB 726 KRDIDINEVEKEEKEEEOETQ---EKGS-----GOEEKGVVINGLDVSPA-----EE 770
OY 172 OKGDK 177
    ::|::|:
DB 771 KKGEDR 776

RESULT 5
YHCS_BACSU
ID YHCS_BACSU STANDARD: PRT: 198 AA.
AC PS4603:
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHEITICAL 22.0 KDA PROTEIN IN CSPB-GLPF INTERGENIC REGION.
GN YHCS.
OS Bacillus subtilis.
OC Bacteria, Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Bacillus.
OC [1]
RN SEQUENCE FROM N.A.
RP STRAIN-168;
RC MEDLINE=97124185; PubMed=8969498;
RX Noback M.A., Terpstra P., Holsappel S., Venema G., Bron S.;
RA "A 22 kb DNA sequence in the cspB-glpPFD region at 75 degrees on the
RT Bacillus subtilis chromosome."
RL Microbiology 142:3021-3026(1996).
CC -----
CC This SWISS-PROT entry is copyrighted. It is produced through a collaboration

```

CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----
 CC EMBL: X96983; CA65703.1; -
 CC EMBL: 299108; CAB12748.1; -
 CC DR SUPPLIST: BG11597; YhcS.
 CC KW Hypothetical protein: Transmembrane.
 CC TRANSMEM 7
 CC SEQUENCE 198 AA: 21976 MW: 2866796CA96A43 CRC64;

Query Match 8.5%; Score 91; DB 1; Length 198;
 Best Local Similarity 23.4%; Pred. No. 1.8; Mismatches 100; Indels 34; Gaps 10;
 Matches 51; Conservative 33;

OY 1 MKWTNRLMTIAGVLLIIVAAFLFAKPHIDNYLHDKDKDEIKDYDKNVK-----EQAS 54
 DB 1 MKKVIPLFIAGLVYAGYGFGLI-----DVTWTKTEQTLKEAKLAKKPEQAS 49
 OY 55 --KDKKQQA-KPQIPKSKSVAGYIEIPDADIKEPVYGPATPEQNLNGVSFAEENESL 110
 DB 50 GTRNSTDAKKNASRPERGQASGILEIPKINAELEPVEG-TPADDLEKGVGHYKDSYYP 108
 OY 111 DQO-NISIAGHFTIDRPNQYFTNLKAAKKGSN-VYEKVGNEFRKYKMTSIDVPTDVG 167
 DB 109 DENGQIVLSGR-----DIVFRRTGELEKGDQLRLISGTEYTELVTKIYDKDDTSII 163
 OY 168 VIDEQKGDQQLTLITCDYDNEKTVGEKKIFVATEV 205
 DB 164 TLOHEK---EELILTYCPFS-YVGNAPKRYIIYGKRV 197

RESULT 6
 XPCPC_XENLA STANDARD; PRT; 1290 AA.
 AC P50532;
 DT 01-OCT-1996 (Rel. 34, Created)
 DT 01-OCT-1996 (Rel. 34, Last sequence update)
 DT 15-JUL-1998 (Rel. 36, Last annotation update)
 DE CHROMOSOME ASSEMBLY PROTEIN XCAP-C.
 GN XCAP-C.
 OS Xenopus laevis (African clawed frog).
 OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
 OC Amphibia; Batrachia; Anura; Mesobatrachia; Pipidoidea; Pipidae;
 OC Xenopodinae; Xenopus.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=95042742; Pubmed=7954811;
 RA Hirano T., Mitchison T.J.;
 RT "A heterodimeric coiled-coil protein required for mitotic chromosome
 RT condensation in vitro."
 RT Cell 79:449-458(1994).
 RL
 CC -I- FUNCTION: REQUIRED FOR BOTH ASSEMBLY AND STRUCTURAL MAINTENANCE OF
 CC MITOTIC CHROMOSOMES.
 CC -I- SUBUNIT: ASSOCIATES WITH XCAP-E PROBABLY AS HETERODIMER.
 CC -I- SUBCELLULAR LOCATION: NUCLEAR. DURING CHROMOSOME ASSEMBLY IN
 CC MITOTIC EXTRACTS, XCAP-C/E WAS RECRUITED TO THE CHROMATIN AND
 CC FORMED A DISCRETE INTERNAL STRUCTURE WITHIN ASSEMBLED CHROMOSOMES.
 CC -I- DOMAIN: CONSISTS OF TWO PUTATIVE CENTRAL COILED-COIL REGIONS
 CC FLANKED BY PUTATIVE GLOBULAR REGIONS AT THE N- AND C-TERMINUS.
 CC -I- SIMILARITY: BELONGS TO THE SMC FAMILY.
 CC -----
 CC This SWISS-PROT entry is copyright. It is produced through a collaboration
 CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
 CC the European Bioinformatics Institute. There are no restrictions on its
 CC use by non-profit institutions as long as its content is in no way
 CC modified and this statement is not removed. Usage by and for commercial
 CC entities requires a license agreement (See <http://www.isb-sib.ch/announce/>
 CC or send an email to license@isb-sib.ch).
 CC -----

CC -----
 DR EMBL: U13673; AA64679.1; -
 KW Mitosis; ATP-binding; Coiled coil; Nuclear protein.
 FT NP_BIND 107 114
 FT DOMAIN 107 114
 FT DOMAIN 264 594
 FT DOMAIN 764 1027
 FT DOMAIN 1094 1129
 FT DOMAIN 1263 1290
 FT DOMAIN 60 65
 FT DOMAIN 747 750
 FT DOMAIN 841 844
 FT DOMAIN 1196 1220
 FT POLY-GLY.
 FT POLY-ALA.
 FT ALA/ASP-RICH (DA-BOX).
 SQ SEQUENCE 1290 AA: 146988 MW: 2931249924FE90F6 CRC64;

Query Match 8.5%; Score 91; DB 1; Length 1290;
 Best Local Similarity 20.3%; Pred. No. 18; Mismatches 60; Indels 38; Gaps 8;
 Matches 36; Conservative 43;

OY 32 YLHD-----KDKD---EKIEQYKKNVKEQA-----SKDKQQAQKQIPKDKSVAGYIE- 77
 DB 320 YIHDIQKRSRKEAQKEKIQEDPKDISKSNLTLETMEKKNALKDKVEKQIKITFTIEE 379
 OY 78 -----IPDADIKEPVYGPATPEQNLNGVSFAEENESLD-QNTSIAGHFTIDRPN 127
 DB 380 NREKFTOLDQVDVDRKRLKSKSKVKRLQK--QLQDKREKVEDELKNVANSQKIIAEET 437
 OY 128 YQFTLKAAKKGSWYFVNGNETRKYKMTSIDR-VKPTDGVLDQKQKDKQQLIT 183
 DB 438 NKKDLERKQ-----EKREKELKNVWDSLKKETQGLQEKKEKELMEIS 483

RESULT 7
 NFM_HUMAN STANDARD; PRT; 915 AA.
 AC P07197;
 DT 01-APR-1988 (Rel. 07, Created)
 DT 01-AUG-1988 (Rel. 08, Last sequence update)
 DT 01-OCT-2000 (Rel. 40, Last annotation update)
 DE NEUROFILAMENT TRIPLET M PROTEIN (160 KDA NEUROFILAMENT PROTEIN)
 DE (NEUROFILAMENT MEDIUM POLYPEPTIDE) (NFM-M) (NEUROFILAMENT 3).
 GN NFM OR NFM OR NEF3.
 OS Homo sapiens (Human).
 OC Eukaryota; Metazoa; Chordata; Vertebrata; Euteleostomi;
 OC Mammalia; Eutheria; Primates; Catarrhini; Hominoidea; Homo.
 RN [1]
 RP SEQUENCE FROM N.A.
 RX MEDLINE=87275853; Pubmed=3608989;
 RA Myers M.W., Lazzarini R.A., Lee V.M.-Y., Schlaepfer W.W., Nelson D.L.;
 RT "The human mid-size neurofilament subunit: a repeated protein
 RT sequence and the relationship of its gene to the intermediate
 RT filament gene family."
 RT EMBO J. 6:1617-1626(1987).
 RL
 RN [2]
 RP SEQUENCE OF ONE OF THE 13 RESIDUE REPEATS.
 RX MEDLINE=88158120; Pubmed=2450354;
 RA Lee V.M.-Y., Ohtos L. Jr., Carden M.J., Hollosi M., Dietzschold B.,
 RA Lazzarini R.A.;
 RT "Identification of the major multiphosphorylation site in mammalian
 RT neurofilaments."
 RL Proc. Natl. Acad. Sci. U.S.A. 85:1998-2002(1988).
 CC -I- FUNCTION: NEUROFILAMENTS USUALLY CONTAIN THREE IF PROTEINS: L, M,
 CC AND H WHICH ARE INVOLVED IN THE MAINTENANCE OF NEURONAL CALIBER.
 CC -I- PTM: THERE ARE A NUMBER OF REPEATS OF THE TRIPLETIDE K-S-P, NFM IS
 CC PHOSPHORYLATED ON A NUMBER OF THE SITES IN THIS MOTIF. IT IS
 CC THOUGHT THAT PHOSPHORYLATION OF NFM RESULTS IN THE FORMATION OF
 CC INTERFILAMENT CROSS BRIDGES THAT ARE IMPORTANT IN THE MAINTENANCE
 CC OF AXONAL CALIBER.
 CC -I- PTM: PHOSPHORYLATION SEEMS TO PLAY A MAJOR ROLE IN THE FUNCTIONING
 CC OF THE LARGER NEUROFILAMENT POLYPEPTIDES (NFM-M AND NFM-H), THE
 CC LEVELS OF PHOSPHORYLATION BEING ALTERED DEVELOPMENTALLY AND
 CC COINCIDENT WITH A CHANGE IN THE NEUROFILAMENT FUNCTION.
 CC -I- SIMILARITY: BELONGS TO THE INTERMEDIATE FILAMENT FAMILY.
 CC -----

```

CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: Y00067; CAA68276.1; -.
DR PIR: A27864; A27864.
DR PIR: A30157; A30157.
DR MIM: 162250; -.
DR INTERPRO: IPR001664; -.
DR PFAM: PF00038; filament. 1.
DR PROSITE: PS00226; IF: 1.
KW Intermediate filament; Heptad repeat pattern; Coiled coil; Neurone;
KW Phosphorylation; Glycoprotein.
FT IN1_MET 0 0
FT DOMAIN 1 103 HEAD.
FT DOMAIN 104 411 ROD.
FT DOMAIN 412 915 TAIL.
FT DOMAIN 104 135 COIL_1A.
FT DOMAIN 136 148 LINKER 1.
FT DOMAIN 149 247 COIL_1B.
FT DOMAIN 248 264 LINKER 12.
FT DOMAIN 265 286 COIL_2A.
FT DOMAIN 287 290 LINKER 2.
FT DOMAIN 291 411 COIL_2B.
FT DOMAIN 613 690 6 X 13 AA TANDEM REPEATS.
FT CARBOHYD 46 46 O-LINKED (GLCNAC) (BY SIMILARITY).
FT CARBOHYD 430 430 O-LINKED (GLCNAC) (BY SIMILARITY).
SQ SEQUENCE 915 AA; 102317 MW; 5F5D3DE34C9D9E50 CRC64;

Query Match 8.4%; Score 90; DB 1; Length 915;
Best Local Similarity 23.5%; Pred. No. 14;
Matches 39; Conservative 26; Mismatches 67; Indels 34; Gaps 6;

QY 26 KPIIDVYLHDKDKDEKIEQYDKNVKQASDKKQQAQKPOIPKSKSVAGYIEIP----- 79
DB 701 KEEEEKVEEVAEPKEEVEK-----KEEKPKDVEPKKKAESPVEEAEEVTTTKSVKXH 755
QY 80 -DADIKPEVYPGPATPEQLNRGVSAEENESLD-----DONISIAHTFIDRPNYQFT 131
DB 756 LEKETEEEGKPLQOQKEKEKAGGSEEGSEKGAKEKKEIDVANGEV-----EGKEEV 811
QY 132 NTKAAKGSNVYKRVGNETRRKYMTSIRDKPPTDVGVLDEQKGDK 177
DB 812 EDETKKESG-----GREEKGVTNGLDLSPA-----DEKKGCGDK 846

RESULT 8
FKB3_YEAST STANDARD; PRT; 411 AA.
ID FKB3_YEAST
AC P38911;
DT 01-FEB-1995 (rel. 31, Created)
DT 01-NOV-1995 (rel. 32, Last sequence update)
DT 15-JUL-1999 (rel. 38, Last annotation update)
DE FK506-BINDING NUCLEAR PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE)
DE (PPIASE) (BC 5.2.1.8) (PROLINE ROTAMASE) (NUCLEOLAR PROLINE
DE ISOMERASE) (FKBP-70).
GN FPR3 OR NP146 OR YML074C.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomycetes.
RN (1)
RN SEQUENCE FROM N.A.
RC STRAIN=YN 214;
RX MEDLINE=95050937; PubMed=7525596;
RA Benton B.M., Zeng J.-H., Thorner J.;
RT "A novel FK506- and rapamycin-binding protein (FPR3 gene product) in
the yeast Saccharomyces cerevisiae is a proline rotamase localized to

```

```

RT the nucleolus."
RL J. Cell Biol. 127:623-639(1994).
RN (2)
RN SEQUENCE FROM N.A.
RP STRAIN=S288C;
RX MEDLINE=94327703; PubMed=8051210;
RA Shan X., Xue Z., Melese T.;
RT "Yeast NP146 encodes a novel prolyl cis-trans isomerase that is
RT located in the nucleolus."
RL J. Cell Biol. 126:853-862(1994).
RN (3)
RN SEQUENCE FROM N.A.
RX MEDLINE=95010686; PubMed=7925954;
RA Manning-Krieg U.C., Henriquez R., Cammas F., Graff P.,
RA Gaveriaux S., Moyva N.R.;
RT "Purification of FKBP-70, a novel immunophilin from Saccharomyces
RT cerevisiae, and cloning of its structural gene, FPR3."
RL FEBS Lett. 352:98-103(1994).
RN (4)
RN SEQUENCE FROM N.A.
RC STRAIN=S288C / AB972;
RA Brown D., Bowman S., Barrell B.G., Rajandream M.A.;
RL Submitted (Oct-1994) to the EMBL/Genbank/DBJ databases.
RN (5)
RN PHOSPHORYLATION OF TYR-184 AND SER-186.
RX MEDLINE=97294697; PubMed=9148902;
RA Wilson L.K., Dhillon N., Thorner J., Martin G.S.;
RT "Casein kinase II catalyzes tyrosine phosphorylation of the yeast
RT nucleolar immunophilin Fpr3."
RL J. Biol. Chem. 272:12961-12967(1997).
CC -I- FUNCTION: PPIASES ACCELERATE THE FOLDING OF PROTEINS. FK506- AND
CC RAPAMYCIN-BINDING PROTEIN. SPECIFICALLY BINDS NUCLEAR LOCALIZATION
CC SEQUENCES. MAY BE INVOLVED IN THE ASSEMBLY OR FOLDING OF RIBOSOMAL
CC PROTEINS.
CC -I- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC
CC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -I- ENZYME REGULATION: INHIBITED BY BOTH FK506 AND RAPAMYCIN.
CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLUS.
CC -I- PTM: PHOSPHORYLATED AT TYROSINE AND DEPHOSPHORYLATED BY THE
CC PHOSPHOTYROSINE-SPECIFIC PHOSPHOPROTEIN PHOSPHATASE PP1.
CC -I- SIMILARITY: BELONGS TO THE FKBP-TYPE PPIASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: U34569; AAB04165.1; -.
DR EMBL: X79379; CAA55924.1; -.
DR EMBL: S73876; AAB31995.1; -.
DR EMBL: Z46373; CAA86504.1; -.
DR PIR: S47927; S47927.
DR PIR: A53544; A53544.
DR HSSP: P20081; 1YAP.
DR SGD: S0004539; NP146.
DR INTERPRO: IPR001179; -.
DR PFAM: PF00254; FKBP; 1.
DR PROSITE: PS00453; FKBP_PPIASE_1; 1.
DR PROSITE: PS00454; FKBP_PPIASE_2; 1.
DR PROSITE: PS00059; FKBP_PPIASE_3; 1.
KW Isomerase; Rotamase; Nuclear protein; Phosphorylation.
FT DOMAIN 60 87 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 88 99 LYS-RICH (HIGHLY BASIC).
FT DOMAIN 101 119 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 173 248 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 250 298 LYS-RICH (HIGHLY ACIDIC).
FT DOMAIN 256 271 NUCLEAR LOCALIZATION SIGNAL (POTENTIAL).
FT DOMAIN 324 411 PPIASE, FKBP-TYPE.
FT MOD_RES 184 184 PHOSPHORYLATION (BY CK2).
FT MOD_RES 186 186 PHOSPHORYLATION (BY CK2).

```

```

FT CONFLICT 122 122 L -> P (IN REF. 1).
FT CONFLICT 240 240 E -> EEE (IN REF. 1).
FT CONFLICT 335 335 L -> F (IN REF. 1).
SO SEQUENCE 411 AA: 46553 MW; A01D24DE0078FE11 CRC64;

Query Match
Best Local Similarity 21.1%; Pred. No. 5.9;
Matches 35; Conservative 37; Mismatches 69; Indels 25; Gaps 6;

OY 30 DNYLHDKDKDEKIEQYDKNVKEDQASKDKKQOAKPOLPKDKSVAGVIEIPADIKEPYYP 89
DB 226 DNDGEDEGEDEEEDEKVEKPEPKSKKKEKKRKEKEEKAK--KKKKVEPKDLEE 283
OY 90 GPATPEQLNGVSPAEENESLDDONISAGHTFIDRPNYQTNLAAKKGWYFK-VG- 147
DB 284 GPTKPKSKKEQDKHKPKSVLEG-GVIEDPRTIGDPQ-----AKGAVGMRYICK 334
OY 148 -----NTRK---YKMTSIRDYKPTDVGVLDEQKDKQOL 181
DB 335 LKNGKVFEDKNTSGKPPAFKLGEGVIAKWDIGVAGMSVGEERRIIT 380

RESULT 9
PMSR_HELPY STANDARD; PRT; 359 AA.
ID PMSR_HELPY
AC 025011;
DT 15-DEC-1998 (Rel. 37, Created)
DT 15-DEC-1998 (Rel. 37, Last sequence update)
DT 30-MAY-2000 (Rel. 39, Last annotation update)
DE PEPTIDE METHIONINE SULFOXIDE REDUCTASE (PEPTIDE MET(O) REDUCTASE).
GN MSRA OR HP0224.
OS Helicobacter pylori (Campylobacter pylori).
OC Bacteria; Proteobacteria; epsilon subdivision; Helicobacter group;
OC Helicobacter.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-26695 / ATCC 700392;
RX MEDLINE-97394467; PubMed-9252185;
RA Tomb J.-F., White O., Kervage A.R., Clayton R.A., Sutton G.G.,
RA Fleischmann R.D., Ketchum K.A., Klenk H.-P., Gill S., Dougherty B.A.,
RA Nelson K., Quackenbush J., Zhou L., Kirkness E.F., Peterson S.,
RA Loftus B., Richardson D., Dodson R., Khaliq H.G., Glodek A.,
RA McEnaney K., Fitzgerald L.M., Lee N., Adams M.D., Hickey E.H.,
RA Berg D.E., Gocayne J.D., Utterback T.R., Peterson J.D., Kelley J.M.,
RA Cotton M.D., Weidman J.M., Fujii C., Bowman C., Watthey L., Wallin E.,
RA Hayes W.S., Borodovsky M., Karp P.D., Smith H.O., Fraser C.M.,
RA Venter J.C.;
RT "The complete genome sequence of the gastric pathogen Helicobacter
RT pylori."
RL Nature 388:539-547(1997).
CC -I- FUNCTION: COULD HAVE AN IMPORTANT FUNCTION AS A REPAIR ENZYME FOR
CC PROTEINS THAT HAVE BEEN INACTIVATED BY OXIDATION. CATALYZES THE
CC REVERSIBLE OXIDATION-REDUCTION OF METHIONINE SULFOXIDE IN
CC PROTEINS TO METHIONINE (BY SIMILARITY).
CC -I- SIMILARITY: TO OTHER BACTERIAL AND EUKARYOTIC PEPTIDE METHIONINE
CC SULFOXIDE REDUCTASES.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: AE000542; AAD07291.1; -
DR TIGR: HP0224; -
DR INTERPRO: IPR002569; -
DR INTERPRO: IPR002579; -
DR PFAM: PF01641; DUF25.1.
DR PFAM: PF01625; PMSR. 1.
KW Oxidoreductase.

```

```

SO SEQUENCE 359 AA: 41275 MW; D481FAC5C60927B3 CRC64;

Query Match
Best Local Similarity 25.1%; Pred. No. 5.4;
Matches 48; Conservative 20; Mismatches 63; Indels 60; Gaps 10;

OY 24 FAKPHIDNYLHD--KDKDEKIEQYDKNVKEDQASKDKKQOAKPOLPKDKSVAGVIEIP-- 79
DB 175 YLKRPSPGCHIDLKKADBEIVDDKRYTKPSPDEVYLLKTKLQYEVQNK---HTEKPE 231
OY 80 -----VDADIEPYYPG-----PATPEQLNGVSPAEENESLDDONIS 116
DB 232 NEYVKEEGEIVDITGTEPLESSADKYDSCGWFSPFSKPIINKDVYKEDDESINRRRIE 291
OY 117 I-----AGHTFIDRPN-----YQPTNLAAKK--GSM-YFKVGNFTR 151
DB 292 VLSRIGKAHLGHVFNDEGKRELGLRVCINSALALRPIPLKDKKEGEGEFTPIYIKKG-ELK 350
OY 152 KYKMTSIRDYK 162
DB 351 KY-----INDKK 357

```

```

RESULT 10
TWIE_CAEEL STANDARD; PRT; 472 AA.
ID TWIE_CAEEL
AC 023525;
DT 15-JUL-1999 (Rel. 38, Created)
DT 15-JUL-1999 (Rel. 38, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE HYPOTHETICAL 54.3 KDA PROTEIN ZK546.14 IN CHROMOSOME II.
GN ZK546.14.
OS Caenorhabditis elegans.
OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabdilitida; Rhabdilioidea;
OC Rhabdilitidae; Peloderinae; Caenorhabditis.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Hallsworth K.;
RL Submitted (JUN-1995) to the EMBL/GenBank/DBJ databases.
CC -I- SUBCELLULAR LOCATION: NUCLEAR; NUCLEOLAR (BY SIMILARITY).
CC -I- SIMILARITY: BELONGS TO THE SURF6 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@sib-sib.ch).
CC -----
DR EMBL: U29380; AAA68745.1; -
DR WORMPEP; ZK546.14; CE02914.
KW Hypothetical protein; Nuclear protein.
SO SEQUENCE 472 AA: 54261 MW; 56EA30387545D3DD CRC64;

```

```

Query Match
Best Local Similarity 8.3%; Score 89; DB 1; Length 472;
Matches 44; Conservative 40; Mismatches 78; Indels 40; Gaps 8;

OY 26 KPHIDNYLHDKDKDEKIEQYD-----KVKEDQASKDKKQOAKPOLPKDKSVAGVIEIPD 80
DB 129 KADSDSEEDSDDEKKEETDEPVAKKOKKESSDDDESDGEPPBNN--GAVEED 185
OY 81 ADIKEPYYPGPATPE-----OLNKGVSFAEENESLDDONISAGHTFIDRPNY 128
DB 186 SDSDEEETPSKPKTYAOSTLKSNGKIDKEIKLEDE--DNESPEIRQIALLRLQK 243
OY 129 QFTNLAAKKG-----SMYFVGNFTRKYKMTSIRDYKPTDVGVLDEQKGD----- 176
DB 244 KLKEKVKERKGGPAKYVSAAEKMAEKKRLKRESKLTKQORRA---EEKKGREAAQV 300

```



```

OY 54 -SKDK-----KQAKPOIP-----KDKSVAGYIEIPDAIKEP 86
DB 1900 VOLDIVSYLNHTNATPIPLFNLYDNEVYVLLPMPSTPKVAFYIKNP----- 1952
OY 87 VYGPATPEQLNRCVSAFENESLDDONISIAHGFIDRPNYQFTNLKAAGKGVYFV 146
DB 1953 -THFVQVEEATNLLISFLHYKFTTDLNOLPPNNTSESTRPGQOTSETLTNKLSEPKR- 2010
OY 147 GNETRKYKMTSIRDVCPFDVGLDEQKQKQTLITCDYN 188
DB 2011 -----KPAVTGLM-----PKSQSIITLSTDTN 2031

RESULT 13
TOP1_DROME STANDARD: PRT; 972 AA.
AC P30189;
DT 01-APR-1993 (Rel. 25, Created)
DT 01-APR-1993 (Rel. 25, Last sequence update)
DT 01-OCT-2000 (Rel. 40, Last annotation update)
DE DNA TOPOISOMERASE I (EC 5.99.1.2).
GN TOP1.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta;
OC Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha;
OC Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RX MEDLINE=93117086; PubMed=1335568;
RA Hsieh T.-S., Brown S.D., Huang P., Postel J.;
RT "Isolation and characterization of a gene encoding DNA topoisomerase
RT I in Drosophila melanogaster."
RL Nucleic Acids Res. 20:6177-6182(1992).
RN [2]
RP SEQUENCE FROM N.A.
RC STRAIN=OREGON-R;
RX MEDLINE=96354910; PubMed=8769417;
RA Zhang C.X., Lee M.P., Chen A.D., Brown S.D., Hsieh T.-S.;
RT "Isolation and characterization of a Drosophila gene essential for
RT early embryonic development and formation of cortical cleavage
RT furrows."
RL J. Cell Biol. 134:923-934(1996).
CC -I- FUNCTION: THE REACTION CATALYZED BY TOPOISOMERASES LEADS TO THE
CC CONVERSION OF ONE TOPOLOGICAL ISOMER OF DNA TO ANOTHER.
CC -I- CATALYTIC ACTIVITY: ATP-INDEPENDENT BREAKAGE OF SINGLE-STRANDED
CC DNA, FOLLOWED BY PASSAGE AND REJOINING.
CC -I- MISCELLANEOUS: EUKARYOTIC TOPOISOMERASE I AND II CAN RELAX BOTH
CC NEGATIVE AND POSITIVE SUPERCOILS, WHEREAS PROKARYOTIC ENZYMES
CC RELAX ONLY NEGATIVE SUPERCOILS.
CC -I- MISCELLANEOUS: WHEN A TOPOISOMERASE TRANSIENTLY BREAKS A DNA
CC BACKBONE BOND, IT SIMULTANEOUSLY FORMS A PROTEIN-DNA LINK, IN
CC WHICH A TYROSYL OXYGEN IN THE ENZYME IS JOINED TO A DNA PHOSPHORUS
CC AT ONE END OF THE ENZYME-SEVERED DNA STRAND.
CC -I- SIMILARITY: BELONGS TO THE EUKARYOTIC TYPE I TOPOISOMERASE FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: M74557; AAA28951.1; -.
DR EMBL: U80064; AAC24158.1; -.
DR PIR: S35521; S35521.
DR HSSP: P11387; 1A35.
DR FLYBASE: FBgn0004924; Top1.
DR INTERPRO: IPR001631; -.
DR PFAM: PF01028; Topoisomerase_I; 1.
DR PRINTS: PR00416; EUPPISMRASEI
DR PROSITE: PS00176; TOPOISOMERASE_I_EUK; 1.
KW Isomerase; topoisomerase; DNA-binding.

```

```

FT DOMAIN 32 39 POLY-HIS.
FT DOMAIN 40 198 SER-RICH.
FT ACT SITE 930 930 DNA CLEAVAGE (BY SIMILARITY).
SQ SEQUENCE 972 AA; 11688 MW; 3764B8BDEEFA30CD CRC64;

Query Match
Best Local Similarity 7.9%; Score 85; DB 1; Length 972;
Matches 41; Conservative 28; Mismatches 86; Indels 26; Gaps 6;

OY 30 DNYLHDKDKDEIEGYDNVKEQASKDKKO-----QAKPQIPKSKYAGYIEIPDAIK 84
DB 140 DKERRDKDKRGSSSSSRHKSSSRDKERSSSSHKSSSSSKSHSSRRHSSSSSK 199
OY 85 E-PVY-----PGPATPEQLNRCVSAFENESLDDONISIAHGFIDRPNYQFTNLKAAG 137
DB 200 DHPYSDGVFAKPEPYSQLMHSGSDVAFQOMQLSGYEAAAGTNTNGNAGNAYKNGY 259
OY 138 KGSWYFVYGVNETRK-----YKMTSIRDVCPFDVGLDEQKQKQTLITCDYN 188
DB 260 EESIYDIKKKEESFNLLSQASSCDYSMSQFADPEPPV-VKHEQSYAREDSMTM-----NYN 314
OY 189 E 189
DB 315 D 315

RESULT 14
YACE_BRELA
ID YACE_BRELA STANDARD: PRT; 195 AA.
AC P56187;
DT 01-NOV-1997 (Rel. 35, Created)
DT 01-NOV-1997 (Rel. 35, Last sequence update)
DT 01-NOV-1997 (Rel. 35, Last annotation update)
DE HYPOTHETICAL PROTEIN IN PTSG 3'REGION (FRAGMENT).
OS Brevibacterium lactofermentum.
OC Bacteria; Firmicutes; Actinobacteria; Actinobacteridae;
OC Actinomycetales; Corynebacterineae.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=ATCC 13869;
RA Yoon K.H.;
RL Submitted (NOV-1993) to the EMBL/GenBank/DBJ databases.
RN [2]
RP IDENTIFICATION.
RA Rudd K.E.;
RL Unpublished observations (JAN-1997).
CC -I- SIMILARITY: BELONGS TO THE UPF0038 FAMILY.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration
CC between the Swiss Institute of Bioinformatics and the EMBL outstation -
CC the European Bioinformatics Institute. There are no restrictions on its
CC use by non-profit institutions as long as its content is in no way
CC modified and this statement is not removed. Usage by and for commercial
CC entities requires a license agreement (See http://www.isb-sib.ch/announce/
CC or send an email to license@isb-sib.ch).
CC -----
DR EMBL: L18875; -. NOT_ANNOTATED_CDS.
DR HSSP: P06143; 1HEY.
DR INTERPRO: IPR001977; -.
DR PFAM: PF01121; UPF0038; 1.
DR PROSITE: PS01294; UPF0038; 1.
KW Hypothetical protein; ATP-binding.
FT NP_BIND 8 15 ATP (POTENTIAL).
FT TER 195 195
SQ SEQUENCE 195 AA; 21135 MW; D7F0333860E0BDD3 CRC64;

Query Match
Best Local Similarity 7.9%; Score 84.5; DB 1; Length 195;
Matches 44; Conservative 21; Mismatches 55; Indels 35; Gaps 10;

OY 65 IPDKSKVAGYIE-----IPDAD--IKEPYVP-PATPEQLNR-GVSAFENESLDDONI 115

```

Db 10 IGGKSTVALLSSEGLIIDLADQIAEIVEPQALAEVAFGEPIIKEDSLDQ-- 67
Qy 116 SIAGHTFIDRPNYQFTNLKAARKGSMYFKVGNET--RKYKMTSLRDYKPTDVGILD---- 170
Db 68 GLAKAFVDEHETALLN-----SIHPRIAETARF---AEAEANGTKVAILYDMLPL 117
Qy 171 EOKGDKQLTLITCDYN-----EKTGVMEK 196
Db 118 VDKGLDRGMDLVLVVDVNVVEERVRLVEKRGLEK 152
RESULT 15
FKB4.DROME STANDARD; PRT; 357 AA.
ID FKB4.DROME STANDARD; PRT; 357 AA.
AC P54397;
DT 01-OCT-1996 (Rel. 34, Created)
DT 01-OCT-1996 (Rel. 34, Last sequence update)
DT 15-JUL-1999 (Rel. 38, Last annotation update)
DE 39 KDA FK506-BINDING NUCLEAR PROTEIN (PEPTIDYL-PROLYL CIS-TRANS ISOMERASE) (PIPIASE) (EC 5.2.1.8).
GN FK506-BP1 OR FKBP39.
OS Drosophila melanogaster (Fruit fly).
OC Eukaryota; Metazoa; Arthropoda; Tracheata; Hexapoda; Insecta; Pterygota; Neoptera; Endopterygota; Diptera; Brachycera; Muscomorpha; Ephydroidea; Drosophilidae; Drosophila.
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-CANTON-S;
RX MEDLINE=95278752; PubMed=7538962;
RA Theopold U., Dal Zotto L., Hultmark D.,
RT "FKBP39, a Drosophila member of a family of proteins that bind the immunosuppressive drug FK506.";
RL Gene 156:247-251(1995).
CC -!- FUNCTION: PIPIASES ACCELERATE THE FOLDING OF PROTEINS. FKBP506 MAY FUNCTION IN A SIGNAL TRANSDUCTION CASCADE DURING EARLY DEVELOPMENT.
CC -!- CATALYTIC ACTIVITY: CIS-TRANS ISOMERIZATION OF PROLINE IMIDIC PEPTIDE BONDS IN OLIGOPEPTIDES.
CC -!- SUBCELLULAR LOCATION: NUCLEAR (PROBABLE).
CC -!- TISSUE SPECIFICITY: UBIQUITOUSLY EXPRESSED, HIGHEST LEVELS IN OVARY.
CC -!- DEVELOPMENTAL STAGE: EXPRESSED DURING ALL STAGES OF DEVELOPMENT WITH HIGHEST EXPRESSION IN EARLY EMBRYO.
CC -!- SIMILARITY: BELONGS TO THE FKBP-TYPE PIPIASE FAMILY.
CC -!- CAUTION: IT IS UNCERTAIN WHETHER MET-1 OR MET-3 IS THE INITIATOR.
CC -----
CC This SWISS-PROT entry is copyright. It is produced through a collaboration between the Swiss Institute of Bioinformatics and the EMBL outstation - the European Bioinformatics Institute. There are no restrictions on its use by non-profit institutions as long as its content is in no way modified and this statement is not removed. Usage by and for commercial entities requires a license agreement (See <http://www.isb-sib.ch/announce/> or send an email to license@isb-sib.ch).
CC -----
CC EMBL: 246894; CA86996.1; -.
DR HSSP: P27124; IROU.
DR FLXBASE: FBgn0013269; FK506-bp1.
DR INTERPRO: IPR001179; -.
DR PRAM: PF00254; FKBP_1.
DR PROSITE: PS00453; FKBP_PIPIASE_1; FALSE_NEG.
DR PROSITE: PS00454; FKBP_PIPIASE_2; 1.
DR PROSITE: PS0059; FKBP_PIPIASE_3; 1.
KM Isomerase; Rotamase; Nuclear protein.
FT DOMAIN 89
FT DOMAIN 119 183 ASP/GLU-RICH (HIGHLY ACIDIC).
FT DOMAIN 186 247 LYS-RICH (BASIC).
FT DOMAIN 269 357 PIPIASE, FKBP-TYPE.
SQ SEQUENCE 357 AA; 39428 MW; F12C6431D4C3B55F CRC64;

Matches 40; Conservative 18; Mismatches 62; Indels 25; Gaps 7;
Qy 47 KNYEQASDKDKQOAKPOIPDKRSKVAGYIEIPDADIKEPYPGPAT--PEQLNRGVSA 104
Db 200 KSGKEQNGVAKKEEAKQO-QKKKEKPEAKKEQPKA--KEPAKQOPASKDPRTITGKIV 256
Qy 105 EE-----NESLDQNISIAHTFIDRPNYQFTNLKAARKGSMYFKVGNETRRYKMTSIR 159
Db 257 DQVVGKGEFAKQKRVSV---YYIGRLQSNKKTFDLSLKCKPKPFALGG-----E 304
Qy 160 DVKPTDVGVLDEQKGRDKQLTLITC 184
Db 305 VIRGMDVGVAGMKVGGRR---VITC 326
Search completed: March 2, 2001, 10:24:33
Job time: 120 sec

Query Match 7.9%; Score 84.5; DB 1; Length 357;
Best Local Similarity 27.6%; Pred. No. 12;

THIS PAGE BLANK (USPTO)

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:20:53 ; Search time 15.51 seconds
(without alignments)
901.840 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKKWNRLMTIAGVLLVA.....YNEKGVWEKKRIFVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 195891 seqs, 67900655 residues

Total number of hits satisfying chosen parameters: 195891

Minimum DB seq length: 0
Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%
Maximum Match 100%

Listing first 45 summaries

Database :
1: pir1:*
2: pir2:*
3: pir3:*
4: pir4:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	104	9.7	352	2 T36719	probable integral
2	102	9.5	552	2 T25593	hypothetical prote
3	101.5	9.4	1089	2 T31583	hypothetical prote
4	97.5	9.1	678	2 A54514	glutamic acid-rich
5	97	9.0	887	2 S59777	hypothetical prote
6	95.5	8.9	644	2 S55395	neurofilament prot
7	94.5	8.8	993	2 S48436	probable RNA helic
8	93.5	8.7	763	2 T08929	hypothetical prote
9	93	8.6	253	2 S73653	hypothetical prote
10	91.5	8.5	312	2 T25894	hypothetical prote
11	91.5	8.5	485	2 S73333	hypothetical prote
12	91.5	8.5	845	2 A45669	Mg260 homolog - My
13	91.5	8.5	1288	2 T46486	neurofilament trip
14	91	8.5	198	2 G69823	chromosomal protei
15	91	8.5	1290	2 A55094	hypothetical prote
16	90	8.4	916	2 A27864	chromosomal protei
17	90	8.4	1192	2 A71623	neurofilament trip
18	89.5	8.3	411	2 S48647	probable secreted
19	89.5	8.3	1609	2 H72013	peptidylprolyl iso
20	89	8.3	359	2 H64547	polymorphic membra
21	89	8.3	359	2 E71960	peptide methionine
22	89	8.3	472	2 T27903	probable peptidase
23	89	8.3	641	2 S41861	hypothetical prote
24	89	8.3	691	2 T32748	gene NS-1 protein
25	89	8.3	1388	2 T17269	hypothetical prote
26	89	8.3	3488	2 T34418	hypothetical prote
27	88.5	8.2	1516	2 E71619	hypothetical prote
28	88	8.2	584	2 A69980	RAD2 endonuclease
29	88	8.2	849	2 S00030	penicillin-binding
					neurofilament trip

30	88	8.2	1291	2 T22382	hypothetical prote
31	87.5	8.1	464	2 T35943	probable hydrolyti
32	87	8.1	605	2 T02350	hypothetical prote
33	87	8.1	712	1 S69782	outer membrane pro
34	87	8.1	1151	2 T24541	hypothetical prote
35	87	8.1	1909	2 A45592	liver stage antigen
36	87	8.1	2078	2 T09326	hypothetical prote
37	86.5	8.0	888	2 T25713	legument protein -
38	85	7.9	507	2 C82901	hypothetical prote
39	85	7.9	608	2 T22897	conserved hypotet
40	85	7.9	627	2 T04562	hypothetical prote
41	85	7.9	665	2 B71609	hypothetical prote
42	85	7.9	972	2 S35521	DNA topoisomerase
43	84.5	7.9	253	2 T36718	probable integral
44	84.5	7.9	357	2 J04090	FK506-binding 39k
45	84.5	7.9	641	2 S41439	gene NS-1 protein

ALIGNMENTS

RESULT 1
T36719
probable integral membrane protein - Streptomyces coelicolor
C:Species: Streptomyces coelicolor
C>Date: 03-Dec-1999 #sequence_revision 03-Dec-1999 #text_change 03-Dec-1999
C:Accession: T36719
R:Murphy, L.; Harris, D.; Bentley, S.D.; Parkhill, J.; Barrell, B.G.; Rajandream, M.A.
submitted to the EMBL Data Library, June 1999
A:Reference number: Z21612
A:Status: preliminary
A:Molecule type: DNA
A:Residues: 1-352 <MUR>
A:Cross-references: EMBL:AL079308; P1DN:CA045217.1; GSPDB:GN00070; SCODEB:SCH69.20C
A:Experimental source: strain A3(2)
A:Genetics:
A:Gene: SCODEB:SCH69.20C

Query Match 9.7%; Score 104; DB 2; Length 352;
Best Local Similarity 23.2%; Pred. No. 0.61;
Matches 47; Conservative 40; Mismatches 82; Indels 34; Gaps 11;

OY 13 GVVLLIIVAALFAKPHIDNYLHDKDKEITEQYDKNVKEQASKDKQQAQPIPKSKV 72
DB 148 GVVLLLFVTVYQLMWNTVRAH-----AQNQAASNDQDWANGKRS-----PGSFEFG 194

OY 73 AGY--TEIPDADIKEPYPCGATPEQLNRGV--SFAEE--NESLDDQ--NISAGHTEI 123
DB 195 QGFALLHPIKLDVVVPIAEISSKVLDRGMVGHVYEDGLKTAMPDAKAGNFGLAGHRT 254

OY 124 DRRNYQFTNLKAAKKSMTYFKVGNETRKTKMYSIRDV-KPTDVGVLDEO-----KGKD 176
DB 255 HGEPEFRYIN--KLEPGGPVIVETQDKYFVVKMASTLPVSPSVSLDVPVKOSGFGKPG 312

OY 177 KQLTLITC-DDYNEKTS--VWEK 196
DB 313 RYTLTCTPEFTSKRYMIVWGK 335

RESULT 2
T25593
hypothetical protein C32E12.1 - Caenorhabditis elegans
C:Species: Caenorhabditis elegans
C>Date: 15-Oct-1999 #sequence_revision 15-Oct-1999 #text_change 15-Oct-1999
C:Accession: T25593
R:Wilson, L.
submitted to the EMBL Data Library, November 1996
A:Description: The sequence of C. elegans cosmid C32E12.
A:Reference number: Z20055
A:Accession: T25593
A:Status: preliminary; translated from GB/EMBL/DBJ

Thu Mar 15 11:53:19 2001

us-09-292-437-3.rpr

Page 6

```

Cy 78 -----IDDAIKBEVYVGPATPOLONGVFAEENSLLD--QWISLACHTFIDRPN 127
      : : : : :
Db 360 NREKFTQLDLDDVDRBRLKHSKSKVKKLX--QLQXDKREKVBELKVPANSOKIIAEET 437
      : : : : :
Cy 128 YQFTNLKAAKAGSNVYFVNGVETKRYKMTSTRD--VRFTDVGVLDEQGRKQKOLLTT 183
      : : : : :
Db 438 NKKDLLEKQK-----EKKEBKIKVMDSLAKETQGLQAEKKIKKIKETLMEIS 483
      : : : : :

```

Search completed: March 2, 2001, 10:22:50
Job time: 117 sec

9

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:22:03 ; Search time 23.18 Seconds

(without alignments)
1041.623 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKKWTNRLMTIAGVILVA.....YNEKTGVMEKRIIVATEVK 206

Scoring table: BLOSUM62
Gap: 10.0 , Gapext 0.5

Searched: 374700 seqs, 117207915 residues

Total number of hits satisfying chosen parameters: 374700

Minimum DB seq length: 0
Maximum DB seq length: 200000000

Post-processing: Minimum Match 08
Maximum Match 1008

Listing first 45 summaries

Database :
1: SP-archaea:*
2: SP-bacteria:*
3: SP-fungi:*
4: SP-human:*
5: SP-invertebrate:*
6: SP-mammal:*
7: SP-mhc:*
8: SP-organellar:*
9: SP-phage:*
10: SP-plant:*
11: SP-rodent:*
12: SP-virus:*
13: SP-vertebrate:*
14: SP-unclassified:*

Pred. No. is the number of results predicted by chance to have a score greater than or equal to the score of the result being printed, and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	DB ID	Description
1	1076	100.0	206	2	09S446
2	184	17.1	193	2	09KBB3
3	182.5	17.0	187	2	09KEV9
4	133.5	12.4	194	2	09KBO9
5	126	11.7	365	2	068213
6	106.5	9.9	231	2	031340
7	104	9.7	352	2	09XA14
8	102	9.5	552	5	P91122
9	101	9.4	1017	10	09S568
10	97.5	9.1	673	5	09U0N1
11	97.5	9.1	810	6	077788
12	97	9.0	887	3	006106
13	96.5	9.0	1147	5	09Y1H3
14	94	8.7	643	5	09U229
15	93.5	8.7	763	10	09S0A1
16	93	8.6	253	2	P75281
17	92	8.6	2268	5	093515
18	91.5	8.5	312	5	P91570
19	91.5	8.5	485	2	P75139

20	91.5	8.5	845	11	063370	063370 ratu
21	91.5	8.5	1202	4	09UNT9	09unt9 homo
22	91.5	8.5	1288	4	095752	095752 homo
23	91.5	8.5	1288	4	09NTJ3	09ntj3 homo
24	91	8.5	954	10	P93826	P93826 arabid
25	90.5	8.4	207	2	09K5S8	09k5s8 bacillu
26	90	8.4	916	4	09UE07	09ue07 homo
27	90	8.4	1192	5	096127	096127 plasmodi
28	89.5	8.3	1609	2	09RB58	09rb58 chlamydia
29	89.5	8.3	1609	2	092605	092605 chlamydia
30	89	8.3	348	10	039448	039448 ciocer
31	89	8.3	359	2	085224	085224 helicobacte
32	89	8.3	359	2	092MK8	092mk8 helicobacte
33	89	8.3	641	12	065023	065023 aleutian m
34	89	8.3	691	5	044741	044741 caenorhabd
35	89	8.3	978	4	075032	075032 homo sapien
36	89	8.3	1388	4	09URF9	09urf9 homo sapien
37	89	8.3	3488	5	P91257	P91257 caenorhabd
38	88.5	8.2	309	2	09R6P5	09r6p5 streptococc
39	88.5	8.2	309	2	P72538	P72538 streptococc
40	88.5	8.2	978	5	0906W5	0906w5 caenorhabd
41	88.5	8.2	1516	5	096154	096154 plasmodium
42	88	8.2	584	2	032032	032032 bacillus su
43	88	8.2	1291	5	09XV10	09xv10 caenorhabd
44	87.5	8.1	241	2	09RPE0	09rpe0 mycoplasma
45	87.5	8.1	309	2	09L5X4	09l5x4 streptococc

ALIGNMENTS

```
RESULT 1
ID 09S446 PRELIMINARY: PRT: 206 AA.
AC 09S446:
DT 01-MAY-2000 (TREMBLrel. 13, Created)
DT 01-MAY-2000 (TREMBLrel. 13, Last sequence update)
DT 01-MAY-2000 (TREMBLrel. 13, Last annotation update)
DE SORTASE.
GN SRTA.
OS Staphylococcus aureus.
OC Bacteria; Firmicutes; Bacillus/Clostridium group;
OC Bacillus/Staphylococcus group; Staphylococcus.
OX NCBI_Taxid=1280;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=8325-4;
RX MEDLINE=99357874; PubMed=10427003;
RA Mazmanian S.K., Liu G., Ton-That H., Schneewind O.;
RT "Staphylococcus aureus sortase, an enzyme that anchors surface
RT proteins to the cell wall."
RL Science 285:760-763(1999).
DR EMBL: AF162687; AAD48437.1;
SQ SEQUENCE 206 AA: 23541 MW: DC3B65C51E145C7B CRC64;
```

Query Match 100.0%; Score 1076; DB 2; Length 206;
Best Local Similarity 100.0%; Pred. No. 4.3e-76;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;

```
QY 1 MKKWTNRLMTIAGVILVAAYLFAPPHIDNYLHDKDKDEKIEQYDKNVEQASRKDKQ 60
DB 1 MKKWTNRLMTIAGVILVAAYLFAPPHIDNYLHDKDKDEKIEQYDKNVEQASRKDKQ 60
QY 61 AKPQIRKSKSVAGYIEIPDADIKVEYVYGPAPPEQLNKGVSFAENESLDDONISIA 120
DB 61 AKPQIRKSKSVAGYIEIPDADIKVEYVYGPAPPEQLNKGVSFAENESLDDONISIA 120
QY 121 TTFDRPNYFTNLKAAKKSMYFKGNETRKXKMTSIRDVKTQDVGVLDEQKGRKQLT 180
DB 121 TTFDRPNYFTNLKAAKKSMYFKGNETRKXKMTSIRDVKTQDVGVLDEQKGRKQLT 180
QY 181 LITCDYNEKTGVMEKRIIVATEVK 206
```


OS *Caenorhabditis elegans*.
OC Eubacteriia; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditidae
OC Rhabditidae; Pelodierinae; *Caenorhabditis*.
OX NCBI_TaxID=6239;
RN [1]
RP
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX MEDLINE=94150718; PubMed=7906398;
RA Wilson S., Almscough R., Anderson K., Baynes C., Berks M.,
RA Bonfield J., Burton J., Connell M., Copey T., Cooper J., Coulson A.
RA Craxton M., Dear S., Du Z., Durbin R., Favallo A., Fulton L.,
RA Gardner A., Green P., Hawkins T., Hillier L., Jier M., Johnston L.,
RA Jones M., Kershaw J., Kirsten J., Laister N., Latreille P.,
RA Lighthouse J., Lloyd C., McMurray A., Mortimore B., O'Callaghan M.,
RA Parsons J., Percy C., Rifkin L., Roopra A., Saunders D., Showkeen I.
RA Smaldon N., Smith A., Sonhammer E., Staden R., Sulston J.,
RA Thierry-Mieg J., Thomas K., Vaudin M., Vaughan K., Watson R.,
RA Watson A., Weinstock L., Wilkinson-Sprat J., Wohlman P.,
RT "2.2 Mb of contiguous nucleotide sequence from chromosome III of C.
RT elegans".
RT [1]
RL Nature 368:32-38(1994).
RN [2]
RP
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RX WILCOX L.;
RN Submitted (JAN-1997) to the EMBL/GenBank/DBJ databases.
RN [3]
RP
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RL Submitted (NOV-1996) to the EMBL/GenBank/DBJ databases.
RN [4]
RP
RC SEQUENCE FROM N.A.
RC STRAIN-BRISTOL N2;
RA Waterston R.;
RN Submitted (MAY-1997) to the EMBL/GenBank/DBJ databases.
RW EMBL; 080032; AAB53878.1; -
SQ RNA-directed DNA polymerase.
SEQUENCE 552 AA; 62427 MW; D570979E3BD7ED94 CRC64;

RESULT	9
Q9SS68	
ID	Q9SS68
AC	Q9SS68:
DT	01-MAY-2000 (TREMBLrel. 13. Created)
DT	01-MAY-2000 (TREMBLrel. 13. Last sequence update)
DT	01-JUN-2000 (TREMBLrel. 14. Last annotation update)
DE	PUTATIVE PHOSPHORIBOSYLTRANSFERINATE TRANSFERASE.
GN	TLJ13.4.
OS	Arabidopsis thaliana (Mouse-ear cress).
OC	Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;

OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
OC Brassicales; Brassicaceae; Arabidopsis.
OX NCBI_TaxID=3702;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN=CV, COLUMBIA;
RA Lin X., Kaul S., Town C.D., Benito M., Creasy T.H., Haas B.,
RA Romling C.M., Koo H., Fujii C.Y., Ultebeck T.R., Barnstead M.E.,
RA Bowman C.L., White O., Nierman W.C., Fraser C.M.;
RT "Arabidopsis thaliana chromosome II BAC T1L13 genomic sequence."
RL submitted (Oct-1999) to the EMBL/GenBank/DBJ databases.
DR EMBL: AC009327; AF03465.1; -
DR INTERPRO: IPR000008; -
DR PFAM: PF00168; C2; 3.
DR PROSITE: PS50004; C2_DOMAIN_2; 3.
DR TRANSFERASE.
SQ SEQUENCE 1017 AA; 114265 MW; ECDB932DB6E279C94 CRC64;

[illegible]

	Query Match	9.1%	Score 97.5;	DB 5;	Length 673;
	Best Local Similarity	24.1%;	Pred. No. 10;		
	Matches	42;	Conservative	32;	Mismatches 73; Indels 27; Gaps 8;
OY	34	HDKDKEKTEIQDYDWVKVEQASRKDDQQAKPOLIPKSKSVAGYTIELPDADIKPEVYPGPAT	93		
		: : : : : : : : :			
Dd	128	HKKKDEKKERKKEDKKRDKRD-KKKKKKKKKKKKKKKKK-----ENSEVMSLYKTGGHKK	180		
OY	94	PEQLNRKGVSFAEEN---ESLDDQNISLSINGHTPIDPNQFTLKAAKAGSMTYEFGVGNET	150		
		: : : : : : : : : : : : : : :			
Dd	181	PK--NATEHGHEINLEYEMWSEINNNAOGGLSSP-VGYREGCGGIISSVH-PTTSNDT	235		

QY 151 RYKMTSIRDVKPTD-----VGLDEQKQKQKQLTTCDDYNEKTYGWEKKR 198
ID 077788 PRELIMINARY; PRT: 810 AA.
AC 077788;
DT 01-NOV-1998 (TREMBlrel. 08, Created)
DT 01-NOV-1998 (TREMBlrel. 08, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE NEUROFILAMENT-M SUBUNIT (FRAGMENT).
GN NF-M.
OS Bos taurus (Bovine).
OC Eukaryota; Metazoa; Chordata; Craniata; Vertebrata; Euteleostomi;
OC Mammalia; Eutheria; Cetartiodactyla; Ruminantia; Pecora; Bovidae;
OC Bovidae; Bovinae; Bos.
NCBI_TaxID=9913;
RN [1]
RP SEQUENCE FROM N.A.
RA Hill W.D., Zhang L., Balin B.J., Sprinkle T.J., Spicer K.,
RT Gearhart D.A.:
RT "The bovine neurofilament M subunit has a novel set of KSP repeats
RT normally restricted to NF-H.";
RL Submitted (SEP-1998) to the EMBL/Genbank/DBJ databases.
CC -I- SIMILARITY: TO ALL OTHER INTERMEDIATE FILAMENT PROTEINS.
DR EMBL: AF091342; AAC36357.1; -
DR INTERPRO: IPR001664; -
DR PFMAM: PF00038; Filament; 1.
DR PROSITE: PS00226; IF; 1.
KW Intermediate filament; Coiled coil; Heptad repeat pattern.
FT NON_TER
SQ SEQUENCE 810 AA; 90799 MW; B8477D85560AC3F6 CRC64;

Query Match 9.1%; Score 97.5; DB 6; Length 810;
Best Local Similarity 23.6%; Pred. No. 13;
Matches 41; Conservative 30; Mismatches 64; Indels 39; Gaps 7;

QY 26 KKHIDNYLHDKRDEKIEQYDKNVKEQASKDK--KQAKPQIPKQSKVAGIETPDADI 83
ID 006106 PRELIMINARY; PRT: 887 AA.
AC 006106;
DT 01-NOV-1996 (TREMBlrel. 01, Created)
DT 01-NOV-1996 (TREMBlrel. 01, Last sequence update)
DT 01-JUN-2000 (TREMBlrel. 14, Last annotation update)
DE HYPOTHELTICAL 101.1 KDA PROTEIN P8283.19.
GN P8283.19.
OS Saccharomyces cerevisiae (Baker's yeast).
OC Eukaryota; Fungi; Ascomycota; Saccharomycetes; Saccharomycetales;
OC Saccharomycetaceae; Saccharomyces.
NCBI_TaxID=4932;
RN [1]
RP SEQUENCE FROM N.A.
RA STRAIN-S288C / AB972;
RA Nelson J.;
RL Submitted (JUL-1995) to the EMBL/Genbank/DBJ databases.
RN [2]

RP SEQUENCE FROM N.A.
RC STRAIN-S288C / AB972;
RA Jia Y., Cherry J.M.;
RL Submitted (SEP-1997) to the EMBL/Genbank/DBJ databases.
DR EMBL: U32445; AAB68082.1; -
DR INTERPRO: IPR000504; -
DR PFMAM: PF00076; rim; 5.
DR PROSITE: PS00030; RNP_1; UNKNOWN_2.
KW Hypothetical protein.
SQ SEQUENCE 887 AA; 101119 MW; 1E34816BD486DEA8 CRC64;

Query Match 9.0%; Score 97; DB 3; Length 887;
Best Local Similarity 18.6%; Pred. No. 16;
Matches 35; Conservative 40; Mismatches 65; Indels 48; Gaps 6;

QY 24 FAKPHIDNVLHDK-----DKDEKIEQYDKNVKEQASKDKKQAKPQIPKDKSVAG 74
ID 0911H3 PRELIMINARY; PRT: 1147 AA.
AC 0911H3;
DT 01-NOV-1999 (TREMBlrel. 12, Created)
DT 01-NOV-1999 (TREMBlrel. 12, Last sequence update)
DT 01-MAY-2000 (TREMBlrel. 13, Last annotation update)
DE POL.
OS Dictyostellium discoideum (slime mold).
OC Eukaryota; Dictyostellida; Dictyostellium.
OC NCBI_TaxID=44689;
RN [1]
RP SEQUENCE FROM N.A.
RC STRAIN-AX4; TRANSPOSON-RETROTRANSPOSON TREE3-C;
RA Szatranski K., Gloeckner G., Dingermann T., Noegel A.A., Eichinger L.,
RT "Non-LTR retrotransposons with unique integration preference
RT downstream of Dictyostellium discoideum transfer RNA genes.";
RL Submitted (MAR-1999) to the EMBL/Genbank/DBJ databases.
DR EMBL: AF134171; AAD43059.1; -
DR INTERPRO: IPR000477; -
DR PFMAM: PF00078; rvc; 1.
SQ SEQUENCE 1147 AA; 134731 MW; D497537E1A024517 CRC64;

Query Match 9.0%; Score 96.5; DB 5; Length 1147;
Best Local Similarity 22.2%; Pred. No. 23;
Matches 46; Conservative 36; Mismatches 90; Indels 35; Gaps 7;

QY 5 TNRUMIAGVNLIVAAVLEFAKPHIDN-VLHD-----KDKDEKIEQYDKNVKEQAS 54
ID 220 TNRETTITTTTKLERLPLWLKEILNNKHIIHGLSELISKNDRKISVEMTK-FKNNVI 278
QY 55 KD--KQQAKPQIPKDKSVAGIETPDADIKEPYVPATPEQLRGVS-FAEENESLD 111
DB 279 RDLKKEQNKIKKKEKKRRYVVIHKLGNSDIT-----FKMKRELNEELSRILEEKKRK 332
QY 112 DONISLAGHTFDIPNYOFTNLKAAGKSMVYFKGNETRKYKMSIRDVKPTDGVLD 171
DB 333 AMDIKLKLHLHOETPSKYLTSLIKSRA-----KDKSIFQIKDKDKNTISD 377

QY 172 QKGDQTLTTCDDYNEKTVGWEKK 198
 Db 378 KENIAKRFVEFYQDOYEKEKDEMEETHK 404

RESULT 14

090229 ID 090229 PRELIMINARY; PRT; 643 AA.
 AC 090229;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-MAY-2000 (TREMblrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE Y56A3A.32 PROTEIN.
 GN Y56A3A.32.
 OS Caenorhabditis elegans.
 OC Eukaryota; Metazoa; Nematoda; Chromadorea; Rhabditida; Rhabditoidea;
 OC Rhabditidae; Peioderinae; Caenorhabditis.
 OX NCBI_TaxId=6239;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Matthews L.;
 RL Submitted (JAN-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA none;
 RT "genome sequence of the nematode C. elegans: A platform for
 investigating biology."
 RL Science 282:2012-2018(1998).
 DR EMBL: AL132860; CAB60511.1; -.
 DR INTERPRO: IPR001100; -.
 DR INTERPRO: IPR001327; -.
 DR PRAM: PF00070; PYR_redox; 1.
 DR PRINTS: PR00368; FADPNR.
 DR PRINTS: PR00411; PNDRTASEI.
 SQ SEQUENCE 643 AA; 71257 MW; EAC5A54980A5F75D CRC64;

Query Match 8.7%; Score 94; DB 5; Length 643;
 Best Local Similarity 28.1%; Pred. No. 18;
 Matches 36; Conservative 12; Mismatches 44; Indels 36; Gaps 6;

QY 34 HDRKDEKIEYDKNVKEQAKSKKQAKPOIPKDKSVAGYIIPDADIEPYPGPA 92
 Db 113 HSKNHEKHHEQKHEE-KEHAPEKKEAPEKPEK-----APEPEPEPAQ 154
 QY 93 ----TPQL--NRGVSPAENESLDDONISLAGH-----TFIDRPNYQFTWLKAAGKSM 141
 Db 155 KQAEQPPQAEKQETKDAEPEQYDDROTAEAVHARRAPAAAEPP-----APSTSKADA 208
 QY 142 YEKVGNE 149
 Db 209 VEKIGEE 216

RESULT 15

09SU01 ID 09SU01 PRELIMINARY; PRT; 763 AA.
 AC 09SU01;
 DT 01-MAY-2000 (TREMblrel. 13, Created)
 DT 01-OCT-2000 (TREMblrel. 13, Last sequence update)
 DE 01-OCT-2000 (TREMblrel. 15, Last annotation update)
 DE HYPOTHETICAL 85.2 KDA PROTEIN.
 GN T15N24.80 OR AT4G2630.
 OS Arabidopsis thaliana (Mouse-ear cress).
 OC Eukaryota; Viridiplantae; Embryophyta; Tracheophyta; Spermatophyta;
 OC Magnoliophyta; eudicotyledons; core eudicots; Rosidae; eurosids II;
 OC Brassicales; Brassicaceae; Arabidopsis.
 OX NCBI_TaxId=3702;
 RN [1]
 RP SEQUENCE FROM N.A.
 RA Bevan M., Zimmermann W., Grueneisen A., Wambutt R., Bancroft I.,
 RA Mewes H.W., Mayer K.F.X., Lemcke K., Scheller C.;

RL Submitted (MAY-1999) to the EMBL/GenBank/DBJ databases.
 RN [2]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (AUG-1999) to the EMBL/GenBank/DBJ databases.
 RN [3]
 RP SEQUENCE FROM N.A.
 RA Zimmermann W., Grueneisen A., Wambutt R., Kalicki J., Wohlmann P.,
 RA Smith A., Mewes H.W., Lemcke K., Mayer K.F.X.;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 RN [4]
 RP SEQUENCE FROM N.A.
 RA EU Arabidopsis sequencing project;
 RL Submitted (MAR-2000) to the EMBL/GenBank/DBJ databases.
 DR EMBL: AL078465; CAB43859.1; -.
 DR EMBL: AL161565; CAB79518.1; -.
 KW Hypothetical protein
 SQ SEQUENCE 763 AA; 85249 MW; D67FAF5BD4DB480B CRC64;

Query Match 8.7%; Score 93.5; DB 10; Length 763;
 Best Local Similarity 24.5%; Pred. No. 24;
 Matches 48; Conservative 27; Mismatches 74; Indels 47; Gaps 8;

QY 25 AKPHIDNYLHD--KDKDEKIEYDKNVKEQAKSD---KKQAKPOIPKDKSVAGYIE 77
 Db 47 AEP--DNWEIDAQIKDKDEKAEDEKSEYVKKKNEDNAETQKMEKVEYTGQGAETNM 104
 QY 78 IPDADIEPYPGPATPEQLNRGVS-----FAENESLDDONISLAGHTFIDRPNYQFTN 132
 Db 105 DEDADGKK-----EQTDGVSVEDTYKNEVESKDN-----NYAKDD 141
 QY 133 LKAAGKSMV---YFKVGNETRRKKMTSIRDVKTLDVG-----VLDEQKGDQTLTI 182
 Db 142 EKETKETDITADHKKAGKEDTQHEADKANCTKGNIGDIKEGTIVDEKGTMDKEVE 201
 QY 183 TCDDYNEKTVGWEKK 198
 Db 202 NGDENKQVNEVEKKEK 217

Search completed: March 2, 2001, 10:23:33
 Job time: 90 sec

GenCore version 4.5
Copyright (c) 1993 - 2000 Compugen Ltd.

OM protein - protein search, using sw model

Run on: March 2, 2001, 10:22:13 ; Search time 43.08 Seconds

(without alignments)
734.072 Million cell updates/sec

Title: US-09-292-437-3

Perfect score: 1076
Sequence: 1 MKKWTNRLMTIAGVILVA.....YNEKGVWEKRIIVATEVK 206

Scoring table: BLOSUM62
Gapop 10.0 , Gapext 0.5

Searched: 957798 seqs, 153513742 residues

Total number of hits satisfying chosen parameters: 957798

Minimum DB seq length: 0

Maximum DB seq length: 2000000000

Post-processing: Minimum Match 0%

Maximum Match 100%
Listing first 45 summaries

Database :

Pending_Patents_AA:*
1: /cgn2_6/ptodata/2/paa/PCTUS.COMB.pep:*
2: /cgn2_6/ptodata/2/paa/US06.COMB.pep:*
3: /cgn2_6/ptodata/2/paa/US07.COMB.pep:*
4: /cgn2_6/ptodata/2/paa/US08.COMB.pep:*
5: /cgn2_6/ptodata/2/paa/US081.COMB.pep:*
6: /cgn2_6/ptodata/2/paa/US082.COMB.pep:*
7: /cgn2_6/ptodata/2/paa/US083.COMB.pep:*
8: /cgn2_6/ptodata/2/paa/US084.COMB.pep:*
9: /cgn2_6/ptodata/2/paa/US085.COMB.pep:*
10: /cgn2_6/ptodata/2/paa/US086.COMB.pep:*
11: /cgn2_6/ptodata/2/paa/US087.COMB.pep:*
12: /cgn2_6/ptodata/2/paa/US088.COMB.pep:*
13: /cgn2_6/ptodata/2/paa/US089.COMB.pep:*
14: /cgn2_6/ptodata/2/paa/US090.COMB.pep:*
15: /cgn2_6/ptodata/2/paa/US091.COMB.pep:*
16: /cgn2_6/ptodata/2/paa/US092.COMB.pep:*
17: /cgn2_6/ptodata/2/paa/US093.COMB.pep:*
18: /cgn2_6/ptodata/2/paa/US094.COMB.pep:*
19: /cgn2_6/ptodata/2/paa/US095.COMB.pep:*
20: /cgn2_6/ptodata/2/paa/US096.COMB.pep:*
21: /cgn2_6/ptodata/2/paa/US097.COMB.pep:*
22: /cgn2_6/ptodata/2/paa/US098.COMB.pep:*
23: /cgn2_6/ptodata/2/paa/US099.COMB.pep:*
24: /cgn2_6/ptodata/2/paa/US100.COMB.pep:*
25: /cgn2_6/ptodata/2/paa/US101.COMB.pep:*
26: /cgn2_6/ptodata/2/paa/US102.COMB.pep:*
27: /cgn2_6/ptodata/2/paa/US103.COMB.pep:*
28: /cgn2_6/ptodata/2/paa/US104.COMB.pep:*

Pred. No. is the number of results predicted by chance to have a
score greater than or equal to the score of the result being printed,
and is derived by analysis of the total score distribution.

SUMMARIES

Result No.	Score	Query Match	Length	ID	Description
1	1076	100.0	206	US-09-292-437-3	Sequence 3, Appli
2	1063	98.8	207	US-08-827-356-5465	Sequence 5465, Ap
3	1063	98.8	207	US-09-611-529-6656	Sequence 6656, Ap
4	1062	98.7	206	US-60-242-578-1023	Sequence 1023, Ap
5	1062	98.7	206	US-60-257-931-3336	Sequence 3336, Ap

6	794.5	73.8	203	18	US-09-450-969-4904	Sequence 4904, Ap
7	160.5	14.9	251	14	US-09-071-035-486	Sequence 486, Ap
8	160.5	14.9	251	15	US-09-134-000-5985	Sequence 5985, Ap
9	158.5	14.7	284	14	US-09-071-035-298	Sequence 298, Ap
10	158.5	14.7	284	20	US-09-634-341-29	Sequence 29, Appl
11	158.5	14.7	284	22	US-60-223-804-29	Sequence 29, Appl
12	158.5	14.7	284	27	US-09-292-437-6	Sequence 6, Appl
13	158.5	14.7	315	15	US-09-134-000-6125	Sequence 6125, Ap
14	157.5	14.6	229	27	US-09-107-532-5472	Sequence 5472, Ap
15	156	14.5	227	27	US-09-292-437-4	Sequence 4, Appl
16	154.5	14.4	197	14	US-09-071-035-488	Sequence 488, App
17	151.5	14.1	325	15	US-09-107-532-5753	Sequence 5753, Ap
18	150	13.9	150	20	US-09-634-341-37	Sequence 37, Appl
19	150	13.9	150	22	US-60-223-804-37	Sequence 37, Appl
20	146	13.6	256	14	US-09-071-035-300	Sequence 300, App
21	143	13.3	283	27	US-09-292-437-34	Sequence 34, Appl
22	140.5	13.1	246	27	US-09-292-437-7	Sequence 7, Appl
23	140	13.0	297	20	US-09-634-341-23	Sequence 23, Appl
24	140	13.0	297	22	US-60-223-804-23	Sequence 23, Appl
25	139.5	13.0	247	19	US-09-583-110-4595	Sequence 4595, Ap
26	139.5	13.0	249	27	US-09-107-433-2710	Sequence 2710, Ap
27	135.5	12.6	296	27	US-09-292-437-35	Sequence 35, Appl
28	129	12.0	248	12	US-08-827-356-3054	Sequence 3054, Ap
29	129	12.0	248	20	US-09-634-341-8	Sequence 3479, Ap
30	126	11.7	283	20	US-09-634-341-8	Sequence 8, Appl
31	126	11.7	283	22	US-60-223-804-8	Sequence 8, Appl
32	126	11.7	365	27	US-09-292-437-5	Sequence 5, Appl
33	125.5	11.7	304	22	US-09-634-341-26	Sequence 26, Appl
34	125.5	11.7	304	22	US-60-223-804-26	Sequence 26, Appl
35	125.5	11.7	304	27	US-09-292-437-36	Sequence 36, Appl
36	125	11.6	348	22	US-09-634-341-32	Sequence 32, Appl
37	125	11.6	348	20	US-60-223-804-32	Sequence 32, Appl
38	123	11.4	251	18	US-60-257-931-2943	Sequence 2943, Ap
39	116	10.8	243	15	US-09-134-000-6184	Sequence 6184, Ap
40	116	10.8	398	15	US-09-107-532-4688	Sequence 4688, Ap
41	115	10.7	236	15	US-09-107-532-6923	Sequence 6923, Ap
42	109.5	10.2	317	15	US-09-107-532-6714	Sequence 6714, Ap
43	108.5	10.1	305	20	US-09-634-341-6	Sequence 6, Appl
44	108.5	10.1	305	22	US-60-223-804-6	Sequence 6, Appl
45	96	8.9	240	20	US-09-634-228-268	Sequence 268, App

ALIGNMENTS

RESULT 1
US-09-292-437-3
; Sequence 3, Application US/09292437
; GENERAL INFORMATION:
; APPLICANT: Olaf Schneewind
; APPLICANT: Sarkis Mazmanian
; APPLICANT: Gwen Liu
; APPLICANT: Hung Ton-That
; TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
; FILE REFERENCE: 510015.213
; CURRENT APPLICATION NUMBER: US/09/292,437
; CURRENT FILING DATE: 1999-04-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 3
; LENGTH: 206
; TYPE: PRT
; ORGANISM: Staphylococcus aureus
US-09-292-437-3

Query Match 100.0%; Score 1076; DB 27; Length 206;
Best Local Similarity 100.0%; Pred. No. 31e-96;
Matches 206; Conservative 0; Mismatches 0; Indels 0; Gaps 0;
OY 1 MKKWTNRLMTIAGVILVAAYLFAPKPHIDNYLHDKDKDEKIQYKKNKEQASKKKQO 60
DB 1 MKKWTNRLMTIAGVILVAAYLFAPKPHIDNYLHDKDKDEKIQYKKNKEQASKKKQO 60

Qy	61	AKPQIPKDSKVAGYIELIPADIDIEPYVPGATPEQJNDRGSAFNEENESLDDQNI	STAGH	120
Db	61	AKQIPKDSKVAGYIELIPADIDIEPYVPGATPEQJNDRGSAFNEENESLDDQNI	STAGH	120
Qy	61	AKQIPKDSKVAGYIELIPADIDIEPYVPGATPEQJNDRGSAFNEENESLDDQNI	STAGH	120
Db	61	AKQIPKDSKVAGYIELIPADIDIEPYVPGATPEQJNDRGSAFNEENESLDDQNI	STAGH	120
Qy	121	TFIDRPYQGFNLNKAARKGSWYVYKGVNENRKYKMSILPDKPDDVGVLEQKQKQ	QOLT	180
Db	121	TFIDRPYQGFNLNKAARKGSWYVYKGVNENRKYKMSILPDKPDDVGVLEQKQKQ	QOLT	180
Qy	181	LITCDYDNEKTGVWEKRRIFAPAEVK	206	
Db	181	LITCDYDNEKTGVWEKRRIFAPAEVK	206	

```

RESULT      2
US-08-827-356-5465
: Sequence 5465, Application US/08827356
: GENERAL INFORMATION:
: APPLICANT: George H. Shimer, Jr.
: APPLICANT: George H. Miller
: APPLICANT: Robert S. Hare
: APPLICANT: Karen J. Shaw
: TITLE OF INVENTION: STAPHYLOCOCCUS AUREUS RELATED
: TITLE OF INVENTION: COMPOSITIONS AND METHODS
: NUMBER OF SEQUENCES: 5574
: CORRESPONDENCE ADDRESSES:
: ADDRESSEE: Schering-Plough Corporation
: STREET: 2000 Galloping Hill Road
: City: Kenilworth
: STATE: New Jersey
: COUNTRY: USA
: ZIP: 07033-0530
: COMPUTER READABLE FORM:
: MEDIUM TYPE: Floppy disk
: COMPUTER: IBM PC compatible
: OPERATING SYSTEM: PC-DOS/MS-DOS
: SOFTWARE: PatentIn
: CURRENT APPLICATION DATA:
: APPLICATION NUMBER: US/08/827,356
: FILING DATE: 01-APR-1997
: PRIOR APPLICATION DATA:
: APPLICATION NUMBER: 60/014,477
: FILING DATE: 01-APR-1996
: APPLICATION NUMBER: 60/016,743
: FILING DATE: 02-MAY-1996
: APPLICATION NUMBER: 60/020,016
: FILING DATE: 14-JUN-1996
: INFORMATION FOR SEQ ID NO: 5465:
: SEQUENCE CHARACTERISTICS:
: LENGTH: 207 amino acids
: TYPE: amino acid
: TOPOLOGY: linear
: MOLECULE TYPE: protein
: HYPOTHEetical: YES
: ORIGINAL SOURCE:
: ORGANISM: Staphylococcus aureus
: FEATURE:
: NAME/KEY: misc_feature
: LOCATION: 1...207
: US-08-827-356-5465

```

98.8%;	Score 1063;	DB 12;	length 207;
Best Local Similarity	99.0%;	Pred. No. 5.7e-95;	
Matches 204;	Conservative 0;	Mismatches 2;	Indels 0;
Gaps 0;			
OY	1	MKKNTNRLMTIACGVLLIIIVAAVLFAFKPHIIDWLVHKKDKDETEQYDKNVKESASDKKKQ	60
	2	MKKNTNRLMTIAGVLLIIIVAAVLFAFKPHIIDLVHKKDKDETEQYDKNVKESASDKKKQ	61
Db			
OY	61	AKQIPKDKSKVAGYIEIPADDIKEPVVPGPATPEQLNRGVSFAEENSLDDONTISAGH	120
	62	AKQIPKDKSKVAGYIEIPADDIKEPVVPGPATPEQLNRGVSFAEENSLDDONTISAGH	121

QY	121	TFIDRPVQOFTNLKAAKGSWYFEKVENRRTKMTSIRDYKPRDVGVLDEQKGDQLT	168
	122	TFIDRPVQOFTNLKAAKGSWYFEKVENRRTKMTSIRDYKPRDVGVLDEQKGDQLT	169
Db	122	TFIDRPVQOFTNLKAAKGSWYFEKVENRRTKMTSIRDYKPRDVGVLDEQKGDQLT	161
QY	181	LITCDYDNEKTGWKKRIRIVATEVK	206
Db	182	LITCDYDNEKTGWKKRIRIVATEVK	207

```

US-09-611-529-6656
RESULT 3
US-09-611-529-6656
Sequence 6656, Application US/09611529
GENERAL INFORMATION:
APPLICANT: George H. Shimer, Jr.
APPLICANT: George H. Miller
APPLICANT: Roberta S. Hare
APPLICANT: Karen J. Shaw
TITLE OF INVENTION: Staphylococcus aureus Related Compositions and Methods
FILE REFERENCE: 1034/IC963051
CURRENT APPLICATION NUMBER: US/09/611,529
CURRENT FILING DATE: 2000-06-30
PRIOR APPLICATION NUMBER: US 09/417,811
PRIOR FILING DATE: 1999-10-14
PRIOR APPLICATION NUMBER: US 09/353,718
PRIOR FILING DATE: 1999-07-14
PRIOR APPLICATION NUMBER: US 09/266,557
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,556
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,555
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,542
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/266,541
PRIOR FILING DATE: 1999-03-11
PRIOR APPLICATION NUMBER: US 09/037,934
PRIOR FILING DATE: 1998-03-10
PRIOR APPLICATION NUMBER: US 09/036,720
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,338
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,334
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,221
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,137
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,082
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,081
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/036,079
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/035,913
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 09/035,744
PRIOR FILING DATE: 1998-03-06
PRIOR APPLICATION NUMBER: US 08/827,336
PRIOR FILING DATE: 1997-04-01
PRIOR APPLICATION NUMBER: US 08/831,156
PRIOR FILING DATE: 1997-04-01
PRIOR APPLICATION NUMBER: US 60/014,477
PRIOR FILING DATE: 1996-04-01
PRIOR APPLICATION NUMBER: US 60/016,743
PRIOR FILING DATE: 1996-05-02
PRIOR APPLICATION NUMBER: US 60/020,016
PRIOR FILING DATE: 1996-06-14
NUMBER OF SEQ ID NOS: 7451
SEQ ID NO 6656
LENGTH: 207
TYPE: PRT
ORGANISM: Staphylococcus aureus

```


US-09-071-035-298

Query Match	14.7%	Score 158.5;	DB 14;	Length 284;
Best Local Similarity	25.3%	Pred. No. 4.6e-07;		
Matches	56;	Conservative	38;	Mismatches 90;
				Indels 37;
				Gaps 7.

[illegible]

```

RESULT 10
US-09-634-341-29
: Sequence 29, Application US/09634341
: GENERAL INFORMATION:
: APPLICANT: Adderson, Elisabeth
: TITLE OF INVENTION: GROUP B STREPTOCOCCUS POYPEPTIDES NUCLEIC ACIDS AND
: FILE REFERENCE: 2511-1-001
: CURRENT APPLICATION NUMBER: US/09/634,341
: CURRENT FILING DATE: 2000-08-08
: NUMBER OF SEQ. ID NOS.: 37
: SOFTWARE: PatentIn Ver. 2.0
: SEQ. ID NO 29
: LENGTH: 284
: TYPE: PR1
: ORGANISM: Enterococcus faecalis
: US-09-634-341-29

```

Query Match	14.7%	Score 158.5;	DB 20;	Length 284;
Best Local Similarity	25.3%	Pred. No. 4.6e-07;		
Matches 56;	Conservative 38;	Mismatches 90;	Indels 37;	Gaps 7.

[illegible]

```

RESULT 11
US-60-223-804-29
; Sequence 29, Application US/60223804
; GENERAL INFORMATION:
; APPLICANT: Aderson, Elisabeth
; APPLICANT: Bohnsack, John
; TITLE OF INVENTION: GROUP B STREPTOCOCCUS POLYPEPTIDES NUCLEIC ACIDS AND
; TITLE OF INVENTION: THERAPEUTIC COMPOSITIONS AND VACCINES THEREOF

```

```

; FILE REFERENCE: 2511-1-001P
; CURRENT APPLICATION NUMBER: US/60/223,804
; CURRENT FILING DATE: 2000-08-08
; NUMBER OF SEQ ID NOS: 37
; SOFTWARE: PatentIn Ver. 2.0
; SEQ ID NO 29
; LENGTH: 284
; TYPE: prt
; ORGANISM: Enterococcus faecalis
; US-60-223-804-29

```

Query Match	14.7%	Score 158.5;	DB 22;	Length 284;
Best Local Similarity	25.3%	Pred. No. 4.6e-07;		
Matches	56;	Conservative	90;	Indels 37;
				Gaps 7;

```

Oy      1  MKKWNRLMTAGTAVVLIVA-----ATLFAPRHIDNLT-----HDKKDE 40
Db      1  MSKSKRRRIIDGEMILLITIGAFAPFVSDALNNLDDQOIIAHYAKASOEKTEMAE 60

Oy      41  KIEQYDNVAKVEQA-----SKDKQOAKPOIPRDKSKVAGYIEIPDADIKVEVP 89
Db      61  LQEKMEKKNGQELAKGSGNPGLDSEFQKTKTKRDKSKYFESHITGVLTIRKINRPLIF 119

Oy      90  GPAPPEQINRGVSPAENE---SLDDONISIAGHTFIDRPNYOFTNLKAAGSMYVYK 146
Db      120  DKTNALLLEKSSSLLEGTSYPTGGTNTHAVISGARGLPQAKL-FTDLPCLKGDEYEV 178

Oy      147  GNETRRYKMTSIRDYKPPDVAVDEQKGRKQKLLITICDDY 187
Db      179  NGKTLAYOVDDIKTVEPTDKDIIIESGD-LVYLLITCTP 218

```

```

RESULT 12
US-09-292-437-6
; Sequence 6, Application US/09292437
; GENERAL INFORMATION:
; APPLICANT: Olaf Schneewind
; APPLICANT: Sarkis Mazmanian
; APPLICANT: Gwen Liu
; APPLICANT: Hung Ton-That
; TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
; FILE REFERENCE: 510015.213
; CURRENT APPLICATION NUMBER: US/09/292.437
; CURRENT FILING DATE: 1999-04-15
; NUMBER OF SEQ ID NOS: 36
; SOFTWARE: FastSeq for Windows Version 3.0
; SEQ ID NO 6
; LENGTH: 284
; TYPE: PRT
; ORGANISM: Enterococcus faecalis
US-09-292-437-6

```

Query Match	14.78;	Score 158.5;	DB 27;	Length 284;
Best Local Similarity	25.38;	Pred. No. 4.6e-07;		
Matches 56; Conservative	38;	Mismatches 90;	Indels 37;	Gaps 7.

```

QY 1 MKKWTNRLMTAGVAVLLIVA-----ALFPAKPHIDNLT-----HDKDKDE 40
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 1 M KSKRRRIIDGFFMILLIIIGAFAPFVSDALNNLTDOOIIAHYQAKASOENTKEMAE 60
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 41 KIEQYDKNVKAEQEA-----SKQKQOAKQPIKDKSKVAGYIEIPDAIDIKREVYP 89
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 61 LQEKMEKKNOELAKKSGNPGLDPSEFOKTKTKKPKDKST FESHITGVLTIRKINRLPFI- 119
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 90 GPATPEOLNRVSVFAEENE---SDDONISIAGHTFIDRPNYQFTNLKAAKSGSMYVFK 146
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 120 DKTALLLEKGSLLLEGTSYPTGGTNNHAVISGRGLPQAKL-FTDLELAKKDEFEIYV 178
      | : : : : : | : : : : : | : : : : : | : : : : : |
QY 147 GNETRKYKMTISRDVKPPDVGVDLEQKQKDKQLTTLITCDY 187
      | : : : : : | : : : : : | : : : : : | : : : : : |
Db 179 NGKTLAYVDQIKTVEPTDTHLIEESGOD-LVTLTLCTPEY 218
      | : : : : : | : : : : : | : : : : : | : : : : : |

```

```

RESULT 13
US-09-134-000-6125
: Sequence 6125, Application US/09134000A
:
: APPLICANT: Lynn Doucette-Stamm et al
:
: TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO ENTEROCOCCUS
:
: FILE REFERENCE: GTC-005
:
: CURRENT APPLICATION NUMBER: US/09/134,000A
:
: CURRENT FILING DATE: 1998-08-13
:
: NUMBER OF SEQ ID NOS: 6810
:
: SEQ ID NO 6125
:
: LENGTH: 315
:
: TYPE: PRT
:
: ORGANISM: Enterococcus faecalis
US-09-134-000-6125

```

Query Match	14.7%	Score 158.5;	DB 15;	Length 315;
Best Local Similarity	25.3%	Pred. No. 5.3e-07;		
Matches	56;	Conservative	38;	Mismatches 90;
				Indels 37;
				Gaps 7;

```

OY      1 MKKMNRLMTAGVLLIYA-----ALYFAKHIDNTL-----HDKKXDE 40
        || : : : : : || : : : : :
Db      32 MKSKKKRRIDIGFMILLITIGAFAPFVSALNNLTDOQIIAHYQAKASQENTKEMAE 91
        || : : : : : || : : : : :
OY      41 KIEQYDKNVEQA-----SKDKQOAKPOLIKRDKSVAGYIEIPDADIKFEEVY 89
        || : : : : : || : : : : :
Db      92 LQEMEKKNQGLAKGNSPGIDPFSEFQTKTKKRDKSYFESHITIGVLTIKINRLPIF- 150
        || : : : : : || : : : : :
OY      90 GPATPEQLNRCVSFAEENE---SLDDONISAGHTTIDRPNYQFTNLKAAKGSMYEYK 146
        || : : : : : || : : : : :
Db      151 DKTALLLEKSSSLEGTSYPTGGTNNHAVISGRGLPQAKL-FTDPELPLKGGDEYIEV 209
        || : : : : : || : : : : :
OY      147 GNETRRYKMTSIROVKPDDVGLDEQKGGKQKLLTLICDDY 187
        || : : : : : || : : : : :
Db      210 NGKTLAYQVDQIKVEPDKIKDLHIESGQ-LVTLITCTCP 249
        || : : : : : || : : : : :

```

RESULT 14
US-09-107-532-5472
; Sequence 5472, Application US/09107532
; GENERAL INFORMATION:
; APPLICANT: Lynn A Doucellette-Stamm and David Bush
; TITLE OF INVENTION: NUCLEIC ACID AND AMINO ACID SEQUENCES RELATING TO
; TITLE OF INVENTION: ENTEROCOCCUS FACIUM FOR
; TITLE OF INVENTION: DIAGNOSTICS AND THERAPEUTICS
; NUMBER OF SEQUENCES: 7308
; CORRESPONDENCE ADDRESS:
; ADDRESSEE: GENOME THERAPEUTICS CORPORATION
; STREET: 100 Beaver Street
; City: Waltham
; STATE: Massachusetts
; COUNTRY: USA
; ZIP: 02354
; COMPUTER READABLE FORM:
; MEDIUM TYPE: CD/ROM ISO9660
; COMPUTER:
; OPERATING SYSTEM:
; SOFTWARE:
; CURRENT APPLICATION DATA:
; APPLICATION NUMBER: US/09/107,532
; FILING DATE:
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/ 085598
; FILING DATE: May 14, 1998
; PRIOR APPLICATION DATA:
; APPLICATION NUMBER: 60/051571
; FILING DATE: July 2, 1997
; ATTORNEY/AGENT INFORMATION:
; NAME: Atinidello, Pamela Deneke

```

1  REGISTRATION NUMBER: 40,489
2  REFERENCE/DOCKET NUMBER: GTC-012
3  TELECOMMUNICATION INFORMATION:
4  TELEPHONE: (781)893-5007
5  TELEFAX: (781)893-8277
6  INFORMATION FOR SHQ ID NO: 5472:
7  SEQUENCE CHARACTERISTICS:
8  LENGTH: 259 amino acids
9  TYPE: amino acid
10  TOPOLOGY: linear
11  MOLECULE TYPE: protein
12  HYPOTHEetical: YES
13  ORIGINAL SOURCE:
14  ORGANISM: Enterococcus faecium
15  FEATURE:
16  NAME/KEY: misc_feature
17  LOCATION: 1...259
18  OS-09-107-532-5472

```

Query Match	14.68;	Score 157.5;	DB 15;	Length 259;
Best Local Similarity	25.48;	Pred. No. 5e-07;		
Matches 57; Conservative	39;	Mismatches 89;	Indels 39;	Gaps 9

```

Qy 1 MKKWTNR--LMTIAGVVLIVLAATLFAKPHINDLHKKDKDETEQYDKNVKEQASDK 58
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 21 MGRWIAEWLLSAGVGLLMLPA-----EASVAKYQAOQOIAAIDRTG 62
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 59 QQAKPOLPKRSKV-----AGIEIIPADIKEPVPGPATPEOLNRSVSAEENESL--- 110
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 63 TAAETDSSLDAKFAELGDPVGLITPISLKLPIYDG-TSKILNENGVLTEGNGDITGG 122
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 111 DDQNISAGHTFIDRPNYQFTNLKAAGKGSWYFRVGNETFRKKYKMTSIRDVKKPTD----- 165
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 122 NGRNPELLIGHSGLYKQDNL-FEDLESYKKGEXEYTIKVDGEQIAVQIDRIEEVQKDLQRF 180
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Qy 166 VGVLDQKGRDKQLTILITCDYINRKTG---VMEKKRIFFVATEVK 206
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :
Db 181 VTYLEPNPNEDNR-VLTMTCTPKGINTHFLVLYGGRVFTTSK 223
    : : : : : : : : : : : : : : : : : : : : : : : : : : : :

```

```

RESULT 15
US-09-292-437-4
: Sequence 4, Application US/09292437
: GENERAL INFORMATION:
: APPLICANT: Olaf Schneewind
: APPLICANT: Sarkis Mazmanian
: APPLICANT: Gwen Liu
: APPLICANT: Hung Ton-That
: TITLE OF INVENTION: IDENTIFICATION OF SORTASE GENE
: FILE REFERENCE: 510015.213
: CURRENT APPLICATION NUMBER: US/09/292,437
: CURRENT FILING DATE: 1999-04-15
: NUMBER OF SEQ ID NOS: 36
: SOFTWARE: FastSeq for Windows Version 3.0
: SEQ ID NO 4
: LENGTH: 227
: TYPE: PRT
: ORGANISM: Streptococcus pyogenes
US-09-292-437-4

```

Query Match	14.5%	Score 156;	DB 27;	Length 227;
Best Local Similarity	32.6%	Pred. No. 5.8e-07;		
Matches 47;	Conservative 19;	Mismatches 60;	Indels 18;	Gaps 7;

QY 56 DKQKQAKPIPKKSKVAGYEIPDADIKFVYGPATPBPOLNKGVSFAENESLDDQNI 115
 Db 32 DKNESLLOI--ENNDIMGVEYVPSIKVLPITY-HTTDEYLTKGAG-----HLPSAL 82
 QY 116 SIAG--HTFIDR----PNTQ-FTNLKAAGKSGVYFKVGNETRRKYKMTSIRDKPKPTVG 167
 Db 83 PVGGDGHCTTIVISAHRGLPSAEMFTNLNIDVKKGDGFYFVRLNKLVLAYVVDQILVYEPQVY 142

OY 168 VLDEQKGDQTLTTCDDYNEXT 191
| | | | | | | | | |
Db 143 SLGVMGKD-YATLVCTPYGVNT 165

Search completed: March 2, 2001, 10:24:18
job time: 125 sec

THIS PAGE BLANK (USPTO)